Advanced personalized gene therapy of B-thalassaemia

Patsali, Petros

Awarding institution:
King's College London

Advanced personalized gene therapy of β-thalassaemia

BY

PETROS PATSALI

A thesis submitted for the degree of Doctor of Philosophy

1st supervisor: Dr Michael Antoniou

Division of Genetics and Molecular Medicine

Department of Medical and Molecular Genetics Guy's Hospital

King's College London

2nd supervisor: Dr Carsten W Lederer

Molecular Genetics Thalassaemia Department

The Cyprus Institute of Neurology and Genetics

KING'S College LONDON

THE CYPRUS INSTITUTE OF NEUROLOGY & GENETICS

Declaration

I, Petros Patsali, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Thalassaemias are amongst the commonest single-gene disorders worldwide, but offer only limited curative treatment choices. Autologous transplantation of gene-therapy-corrected cells is therefore investigated by numerous groups and is already under clinical trials for β-thalassaemia, based on gene addition by lentiviral vectors (LVs).

The main aim of this project was the development of personalised gene therapy (GT) for β-thalassaemia caused by the common and severe *HBB*$^{IVSI-110}$ mutation, which results in missplicing of intron 1. The resulting aberrant mRNA interferes with protein expression from normal endogenous or vector-encoded *HBB*. The two therapy approaches taken were a) modification or co-transduction of the LV MA821_HBB ("GLOBE") to achieve concurrent vector-derived *HBB* expression and RNAi-based suppression of aberrant RNA and b) development of designer nucleases for genome editing and functional correction of the *HBB*$^{IVSI-110}$ mutation.

To establish proof of principle for both approaches we developed transgenic murine erythroleukaemia cell lines MEL MA821-HBB$^{IVSI-110}$ and MEL MA821-HBB^{Normal} with on average two vector copies per genome (VCN) and an additional, clonal cell line of VCN=1 for MEL MA821 HBB$^{IVSI-110}$.

Approach a)

Depletion of aberrant HBB$^{IVSI-110}$ mRNA might enhance translation of vector-derived or residual endogenous HBB mRNA and could be achieved by HBB$^{IVSI-110}$-specific RNAi, mediated by expression of lentivirally delivered shRNAs.

To this end, four different HBB$^{IVSI-110}$ mRNA-specific shRNAs were cloned into the U6-promoter-driven pLKO.I lentiviral vector (LV). In clonal MEL HBB$^{IVSI-110}$ cells, two constructs, pLKO.I shIVSI-110 Mid & Mid2, exhibited no discernible change of normal:aberrant HBB mRNA but a significant increase of HBB chains relative to untransduced samples. Importantly, our approach was then validated by transduction of HBB $^{IVSI-110}$-patient-derived hCD34^{+} cells, in which we observed a ~30% increase of HBB:HBA chain ratios in samples transduced with shIVSI-110 MID, regardless of co-transduction with other vectors. These data indicate that RNAi-targeting of aberrant HBB mRNAs, if not therapeutic in its own right, could substantially improve efficiency of GT by *HBB* gene augmentation and could thus lower conditioning, VCN and gene-expression requirements for LV-based gene therapies.

Approach b)

As an alternative to gene augmentation we established a novel genome-editing approach in which permanent functional correction the $HBB^{IVSI-110}$ mutation was achieved with the targeted disruption of the aberrant SA site and/or its sequence context. This approach, based on the efficient non-homologous end-joining (NHEJ) repair pathway, used $HBB^{IVSI-110}$ - specific designer nucleases, transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 RNA-guided endonuclease (RGENs).

Assessment of targeted disruption efficiency of our designer nucleases on gDNA or episomal reporter plasmids indicated the superiority of the HBB TALEN over RGEN for $HBB^{IVSI-110}$.

Plasmid transfection of $HBB^{IVSI-110}$ specific - designer nucleases into MEL MA821-$HBB^{IVSI-110}$ transgenic cells resulted in functional correction at the RNA (RT-qPCR) and protein level (Immunoblots). In addition, full characterisation (type, frequency and context) of induced insertions/deletions (INDELs) was achieved by T7 Endonuclease 1 assay, decomposition of sequence traces (TIDE) and Sanger sequencing of TOPO clones from edited bulk cells. Using edited MEL MA821-$HBB^{IVSI-110}$ clones we moreover correlated specific INDELs with *HBB* expression at the RNA and protein level, which confirmed our hypothesis that functional correction of splicing could be achieved by disruption of upstream sequences of the aberrant SA site, leaving intact the $HBB^{IVSI-110}$ mutation.

Subsequent validation of our nucleases in patient-derived $CD34^+$ cells by electroporation of HBB TALEN L1/R1 plasmid DNA was impaired by low targeted disruption (12%) despite high-transfection efficiencies. Even though a marginal increase in the HBB:HBA chain ratio was detected, analyses of mRNA levels were inconclusive, calling for additional experimentation in $CD34^+$ cells with more appropriate expression systems.

Overall, proof of principle was established for NHEJ-based functional repair of $HBB^{IVSI-110}$, and we suggest that the approach can be adapted for the functional correction of other of disease-causing gain-of-function mutations in non-coding regulatory regions.

Acknowledgements

Firstly, I would like to thank my PhD supervisors Dr Michael Antoniou and Dr Carsten W Lederer for their encouragement and guidance through all the years of my PhD project. Many thanks also go to the Dr Marina Kleanthous, for introducing me to the field of thalassaemias and accepting me in Molecular Genetics Thalassaemia Department at the Cyprus Institute of Neurology and Genetics.

In addition, I would also like to thank Prof Toni Cathomen and Dr Claudio Mussolino for giving me the opportunity to work in their departments at the University of Freiburg for the development of genome editing tools, which have turned out critical components of major results sections of my PhD work. Of course I would also like to thank all the people in their departments for their help and hospitality during my staying in Freiburg.

Furthermore, I would like to thank the PhD students and researchers of the Molecular Genetics of Thalassaemia Department, Coralea Stephanou, Constantinos Loucari, Pavlos Fanis, Andria Theodorou, Stella Voskou, Stephanie Byrou, Petros Kountouris and Marios Phylactides for their support and advices. I would like to acknowledge the contribution of Ioanna Christodoulou, Petros Ladas, Argyro Floga and especially Panayiota Papasavva for their assistance in different parts of the PhD project, as specified in the respective results sections.

My thanks also go out for the support and hospitality I received from the all remaining members of the Molecular Genetics Thalassaemia Department, George Christopoulos, Thessalia Papasavva, Eleni Karitzie, Eleni Pavlou, Miranda Petrou and Xenia Feleki. I gratefully acknowledge the funding received from the Research Promotion Foundation of Cyprus that enabled the commencement of this project.

Finally, I would I dedicate my PhD thesis to my family as an acknowledgement for their support and understanding through all these years.

Abbreviations

The table list only abbreviations that are not included to the list of acceptable abbreviation in articles defined by the journal of biological research, as also not genes beside murine and human globin genes.

Abbreviation	Meaning
2,3-BPG	2,3-bisphosphoglycerate
A	adenine
AAV	adeno-associated virus
ACH	active chromatin hub
AGM	aorta-gonad-mesonephros
AGO	Argonaute protein
allo-HSCT	allogeneic haematopoietic stem cell transplantation
Amp^R	ampicillin resistance
APS	ammonium persulfate
ASHP	α-haemoglobin-stabilizing protein
ASON	anti-sense oligonucleotides
asRNA	anti-sense RNA
Baso	basophilic erythroblast
BFU-E	burst forming unit –erythroid
BM	bone marrow
BMT	bone marrow transplantation
BPS	branch point site
C	Cytosine
Cas9	CRISPR-associate protein
cDNA	complementary DNA
CFC	colony-forming cell
CFU-E	colony-forming unit-erythroid
CFU-G	colony-forming unit-granulocyte
CFU-GEMM	colony-forming unit-granulocyte, erythroid, macrophage, megakaryocyte
CFU-GM	colony-forming unit-granulocyte, macrophage
CFU-M	colony-forming unit-macrophage
CFU-S	colony-forming unit spleen chicken
cHS4	chicken hypersensitive site-4
CLP	common lymphoid progenitor
CMP	common myeloid progenitor
cPPT	central polypurine tract
CRISPR	clustered regularly interspaced short palindromic repeats
C_t	threshold cycle
DECID	decay-induced complex
dH2O	distilled water
dNTP	deoxyribonucleotide triphosphate
DSB	double strand breaks
dsDNA	double stranded DNA double-stranded

E. coli	*Escherichia coli*
EJC	exon junction complex
EMH	extramedullary haematopoiesis
EPO	Erythropoietin
EryD	definitive erythroid
EryP	primitive erythroid
ESEs and ISEs	exonic and intronic enhancers, respectively
ESSs and ISSs	exonic and intronic silencers, , respectively
FDA	U.S. Food and Drug Administration
FL	foetal liver
FW	denoting forward primer
G	Guanine
G-CSF	granulocyte colony-stimulating factor
gDNA	genomic DNA
GMP	granulocyte-monocyte progenitor
GP	granulocyte progenitor
GPA	glycophorin A
gRNA	guide RNA
GT	gene therapy
GvHD	graft versus host disease
HBA	human α-globin genes
HbA1	adult haemoglobin 1 ($\alpha_2\beta_2$)
HbA2	adult haemoglobin 2 ($\alpha_2\delta_2$)
HBB	human β-globin gene
Hbb1	murine -β major globin gene
Hbb2	murine -β major globin gene
$Hbb^{th1/th1}$	murine thalassaemic model
$Hbb^{th3/+}$	murine thalassaemic model
HBD	human δ-globin gene
HbE	embryonic haemoglobin
HBE	Human ε-globin gene
Hbe/εy	murine εy-globin gene
HbF	foetal haemoglobin
HBG	human γ-globin genes
HBG1	human Aγ-globin gene
HBG2	human Gγ-globin gene
HbH	β_4 tetramers
HBS	HEPES-buffered saline
HBZ	human ζ-globin gene
HDR	homology-directed repair
HEK293T	human embryonic kidney 293T human
HEL	human erythroleukaemia
heterogeneous nuclear ribonucleoprotein	hrRNP
HGVS	human genome variation society

HPFH	hereditary persistent foetal haemoglobin
HR	homologous recombination
HS	hypersensitive element
HSC	haematopoietic stem cell
HU	hydroxyurea
IDLV	Integrase deficient lentiviral vectors
IMDM	Iscove's Modified Dulbecco' Medium
INDELs	insertions or deletions
iPSCs	induced pluripotent stem cells
IVS	intervening sequence
Kan^R	kanamycin resistance
kb	kilobases
LB	lysogeny broth
LMPP	lymphoid-primed multipotential progenitor
LT-HSC	long-term-repopulating haematopoietic stem cells
LV	lentiviral vector
M	molar
m	milli (10^3)
MACS	magnetic-activated cell sorting
MARE	MAF-recognition element
MEL	murine erythroleukaemia cells
MEP	megakaryocyte-erythrocyte progenitor
MFI	mean fluorescence intensity
MGB	minor groove binding
miRNA	microRNA
MOI	multiplicity of infection
MPP	multipotent progenitor
mRNP	messenger ribonucleoprotein
multispecies conserved sequence	MCS
n	nano (10^{-9})
NHEJ	non-homologous end joining
NK	natural killer
NLS	nuclear localisation signal
NMD	non-sense-mediated decay
NOD-SCID	non-obese diabetic/severe combined immunodeficiency NOD-scid
NTC	non-template control
Ortho	orthochromatophilic erythroblasts
P/S	penicillin/streptomycin
PABP	poly (A) binding protein
PAM	protospacer-adjacent motif
PB	peripheral blood
PBMC	peripheral blood mononuclear cells
PEI	polyethylenimine
PGK	phosphoglycerate kinase
Poly	polychromatophilic erythroblasts

PRE	pyrimidine-rich element
PreproEr	Pre-proerythroblast
Pri-miRNA	primary microRNA
ProEr	Proerythroblast
PS	Phosphatidylserine
PTC	pre-termination codons
PYT	poly-pyrimidine tract
qPCR	quantitative polymerase chain reaction
R	haemoglobin relaxed state
RGEN	RNA-guide endonuclease (aka CRISPR/Cas9)
RIPA	radioimmunoprecipitation assay buffer
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNA pol	RNA polymerase
ROS	reactive oxygen species
RP-HPLC	reversed-phase high performance liquid chromatography
RPMI	Roswell Park Memorial Institute medium
RRE	Rev-responsive element
RSV	Rous sarcoma virus
RTemp	room temperature
RT-qPCR	reverse-transcription quantitative polymerase chain reaction
RV	denoting reverse primer
RVDs	repeat-variable di-residues
SA	splice acceptor
SCD	sickle cell disease
SCF	stem cell factor
SD	splice donor
SFFV	spleen focus-forming virus
shIVSI-110_Down	shRNA targeting downstream half of IVSI-110 HBB aberrant mRNA
shIVSI-110_Mid	shRNA targeting central part of IVSI-110 HBB aberrant mRNA
shIVSI-110_Mid2	2nd shRNA targeting central part of IVSI-110 HBB aberrant mRNA
shIVSI-110_Up	shRNA targeting upstream half of IVSI-110 HBB aberrant mRNA
shRNA-miR	microRNA-adapted shRNA
SIN	self-inactivating design
snRNA	small nuclear RNA
snRNPs	small nuclear ribonucleoprotein
SR	serine/arginine-rich protein
ST-HSC	short-term haematopoietic stem cells
SURF	surveillance complex
T	thymine // haemoglobin tensed state
T87Q	codon 87 change from threonine to glutamine in HBB
TALENs	transcription activator-like effector nucleases
TBE	tris-borate-EDTA
TEMED	tetramethylethylenediamine
TIDE	tracking of INDELs by decomposition
TREX2	three-prime repair exonuclease 2

TU	transduction units
UPF	up-frameshift factor
V	Volts
VCN	vector copy number
VSVG	vesicular stomatitis virus glycoprotein G
WHO	World Health Organisation
WPRE	woodchuck hepatitis virus posttranscriptional regulatory element
YS	yolk sac
ZF	zinc-fingers
ZFN	zinc-finger nucleases
αMEM	α-Minimum Essential Medium
βLCR	β-locus control region
βp	β-globin promoter
Δ	denoting deletion
μ	micro (10^{-6})
Ψ	lentiviral packaging signal // pseudogene

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Haematopoiesis

Blood is a specialized tissue responsible for the delivery of important substances, such as nutrients and oxygen, to cells and at the same time is essential for the removal of metabolic waste and carbon dioxide (CO_2) away from those same cells. In addition, blood plays a crucial role in the immune defence of the body against infections. The human blood system is ~7 % of the total body mass, which corresponds to approximately 4–5 L in adults. 45% of blood is constituted by blood cells and the remainder by blood plasma, which is mostly water (92%), proteins, hormones, mineral ions and glucose. The major plasma proteins are albumin, whose primary role is to maintain colloid osmotic pressure, and fibrinogen, which is essential for the formation of blood clots and globulins (including antibodies). Plasma is essential for the maintenance of body temperature, blood volume and pressure, circulation of nutrients such as amino acids, fatty acids and glucose, supplying proteins for immunity and clotting and preserving normal pH balance (7.35–7.45) in the body [1].

The cellular compartment of the blood includes the highly abundant red blood cells (RBCs) occupying 44% of total blood volume. Mammalian RBCs are enucleated biconcave cells without any organelles and filled with haemoglobin (Hb) molecules. The main role of RBCs is the transportation of oxygen (O_2) from lungs to the rest of the body and removal of the CO_2. Other blood cells, although far less numerous than RBCs, have vital biological roles. Platelets are formed with the cytoplasmic degradation of megakaryocytes and are essential for the formation of blood clots upon injuries. Lymphocytes are divided in B- and T-lymphocytes and natural killer (NK) cells, which constitute the main defences of the organism against foreign material and infections [1]. They either directly recognize and destroy (through phagocytosis) infected cells (T cells and NK cells) or produce antibodies (in the case of B-cell derived plasma cells), i.e. immunoglobulins specific for foreign antigens and acting as signal for the elimination of foreign objects. Granulocytes are divided in three populations, neutrophils, eosinophils and basophils, and are key mediators of the innate immune system (phagocytosis and inflammatory response). Blood monocytes may enter damaged tissues through the endothelium of blood vessels, in order to differentiate to macrophages, which are responsible for eliminating dead cells, cell debris and foreign substances by phagocytosis. Dendritic cells are specialized antigen-presenting cells. They are derived from monocytes, and their main role is the phagocytosis and stimulation of immune responses via antigen presentation to lymphocytes. Finally, mast progenitors migrate to different body sites where they differentiate to mature mast cells, which are implicated in wound healing, in the protection against pathogens and in allergy and anaphylaxis [1].

1.1.1 *Hierarchical ontogeny*

The concept of haematopoietic stem cells was first published in the 1960s by Till and McCulloch, who showed with mouse syngeneic bone marrow transplantations (BMT) in lethally irradiated mice the formation of donor-derived colonies in the host spleen. After analysis of these colonies, also known as colony forming units-spleen (CFU-S), they discovered that a subset of bone marrow (BM) cells had the ability to self-replicate and to give rise to multiple cell types from myeloerythroid lineages. In addition, these colonies were able to reform CFU-S in secondary transplanted mice [2]. This pioneering study thus defined two major attributes that characterize haematopoietic stem cells (HSCs), a) self-renewal (in every cell division at least one of the two daughter cells is an identical undifferentiated HSC) and b) multi-potency (differentiation of cells to all blood lineages). HSCs have three potential fates, 1) self-renewal (preserving the HSC population), 2) differentiation and 3) apoptosis. Self-renewing divisions of HSC can be either symmetric, in which two identical daughter HSCs are produced increasing the number of the HSC repertoire, or asymmetric, in which one of the daughter cells remains an undifferentiated HSC and the second is differentiation or undergoes apoptosis. The ability of HSCs to expand (by symmetric divisions), undergo apoptosis and maintain the population (by asymmetric divisions) is essential to maintaining normal haematopoiesis throughout adult life [1,3]. The ten distinct mature cell types pointed out in section 1.1 (RBC, macrophages/monocytes, megakaryocytes, granulocytes, T- and B- lymphocytes, NKs, mast cells and dendritic cells), indicate the multi-lineage capacity of HSCs. HSC and early progenitors can be characterized by flow cytometry based on their cell surface markers and their ability to efflux fluorescence dyes [1,4].

A proposed hierarchical model of haematopoietic development, during which the multi-lineage capacity is progressively lost, is shown in Figure 1. There are three subsets of HSCs: 1) the long-term reconstituting HSC (LT-HSCs), which are able to self-renew and are found at relatively low frequencies (1:10000) in the BM, 2) the short-term reconstituting HSCs (ST-HSCs), whose self-renewal capacity is limited, with relative frequency of 1:2000 in the BM, and finally 3) the multi-potent progenitors (MPP) which they lose their self-renewal capacity but maintain their multi-potency [1]. MPP cells differentiate to more lineage-restricted progenitors (oligo-potent progenitors), the common lymphoid progenitor (CLP) and the common myeloid progenitor (CMP). CMP oligo-potent progenitor gives rise to more lineage-restricted oligo-potent progenitors the megakaryocyte/erythrocyte progenitors (MEP) and granulocyte/macrophage progenitors (GMP). Subsequent differentiation leads to the rise of uni-potent progenitors which are lineage-restricted and eventually differentiate to all the mature blood cells described above. Ambiguities in the existing

understanding of cell markers are still common, for example in the case of dendritic progenitors, which carry cell markers CD8α^{+}, CD8α^{-} and plasmacytoid, that can be derived either from CLP or CMP [4]. Additional work is therefore still required to establish and clarify the hierarchical model of haemotopoiesis in detail, also towards defining the ideal surface marker (combination) for the selection of LT-HSC.

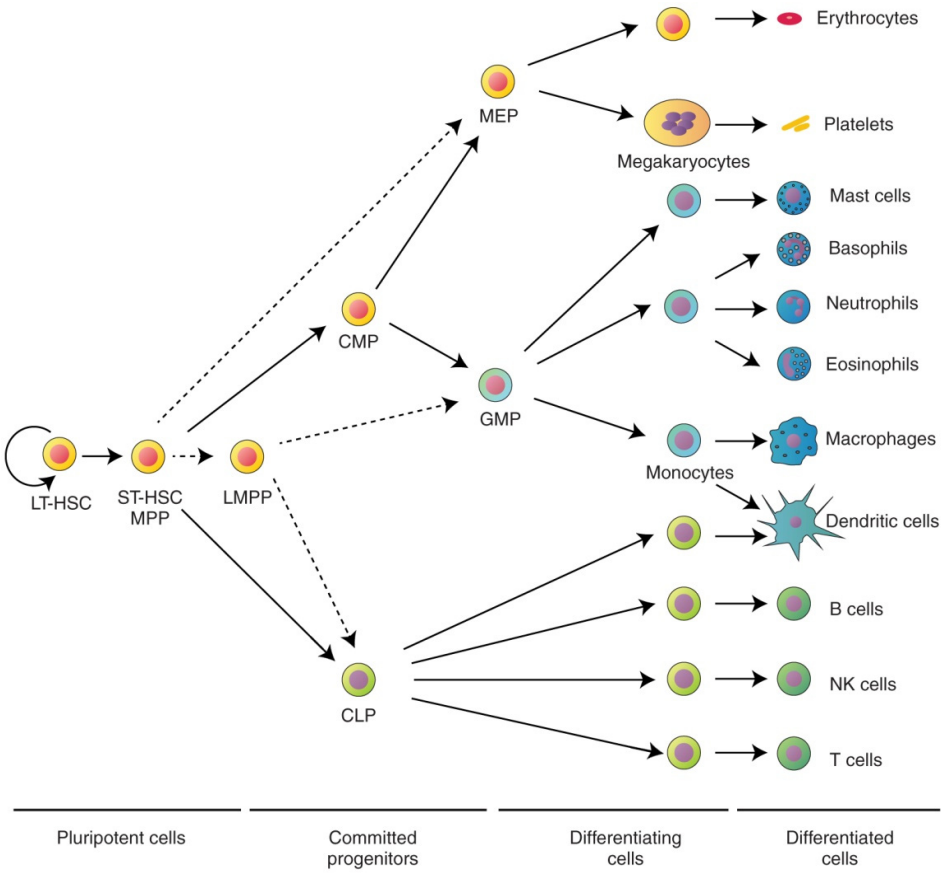


Figure 1: Proposed models for the haematopoietic hierarchy.
There are two overlapping proposed models. The first model (solid arrows) proposed by the Weissman group suggests that multipotent progenitors (MPPs or short-term HSCs (ST-HSCs) are differentiated to committed progenitors, either a common lymphocyte progenitor (CLP) or a common myeloid progenitor (CMP), which then differentiated to either a granulocyte-macrophage progenitor (GMP) or a megakaryocyte-erythroid progenitor (MEP). The second model, proposed by the Jacobson group (dashed arrows), suggests that a multipotent progenitor (MPP or ST-HSC) differentiated directly to MEPs and to lymphoid-primed multipotent progenitor (LMPPs), which have the potential to give rise to both CLPs and GMPs. LT-HSC, long-term haematopoietic stem cell; NK cell, natural killer cell. Figure reproduced from [5].

1.2 Erythropoiesis

RBCs are the most abundant cell population in blood, and their primary goal is to transfer O_2 from the lungs to the body and to remove CO_2 in the reverse direction. Critical to this task is the main RBC constituent, Hb. Terminally differentiated mammalian RBCs circulate in the blood stream enucleated, which in comparison to other higher animals is a unique characteristic of mammals [6]. In

general, RBC life-time is approximately 120 days, therefore a steady-state production of RBC (~2 x 10^6 cell/second) is required to maintain RBC mass at normal levels [7]. Blood homeostasis is achieved through a multi-stage process called erythropoiesis, in which haematopoietic progenitors follow a differentiation pathway to mature RBCs, characterised by a reduction of proliferation rate, morphological changes (e.g. size reduction, nuclear condensation, loss of cytoplasm and organelles, enucleation) and increase of Hb content [8]. Erythropoiesis is regulated by transcription factors, such as GATA1, KLF1, SCL/TAL1 and Lmo2), and by microRNA (miRNAs) and chromatin modification [9,10]. Erythropoiesis is the production of RBCs, which is regulated based on the developmental-stage-specific erythropoietic tissues and globin genes and is divided into primitive (EryP) and definitive (EryD) erythropoiesis [11]. Survival, self-renewal and differentiation of HSCs are depended on their microenvironment, which includes interaction with each other or with the neighbouring cells and secreted factors. Anatomical localisation of these micro-environments, known as niches, changes based on the stage of ontogeny and development, also dependent on properties of the accommodated HSC.

1.3 Primitive erythropoiesis

Primitive erythropoiesis is initiated by mesodermal precursor cells, so-called haemangioblasts, after their migration to the yolk sac (YS) [12]. At that point, pools of immature erythroid primitive progenitor colony forming cells (EryP-CFC), which are enveloped by endothelial cells, form blood islands. EryP-CFC are large blast-like cells and are identified by their ability to form colonies in semisolid methylcellulose [11]. In mouse, it has been shown that EryP-CFC expand in number for 48 hours in the YS before disappearing, indicating the transient nature of primitive erythropoiesis [12]. As a unique characteristic of primitive erythropoiesis, primitive erythroblasts enter circulation as nucleated cells, with the onset of cardiac contractions. There they expand in number in several cell divisions, before starting a semi-synchronous differentiation to mature primitive erythrocytes. EryP erythroblasts are very large cells that experience morphological changes as they differentiate into more mature cells, including a decrease in cell size, nuclear condensation, and Hb accumulation reaching to steady state levels of 80–100 pg/cells [12]. For a long time it was believed that mature EryP erythrocytes retain their nucleus, suggesting a similarity to the nucleated RBCs of birds, fishes and amphibians, rather than to the mammalian foetal and adult definitive enucleated RBC. A study on mouse embryo by K. E. McGrath [13] indicated that EryP erythroblasts enucleate by nuclear extrusion, and suggested that this occurs in the foetal liver (FL) in association with macrophages, leading to the formation of enucleated EryP erythrocytes and a transient population of extruded nuclei, also known as EryP pyrenocytes (Mouse E 14.5-16.5) [13]. EryP erythrocytes remain in blood circulation for 4–5 days post-

enucleation, covering Hb requirement of the embryo until foetal and post-natal definitive erythropoiesis in BM are established. Human EryP cells express a range of embryonic and adult globin chains, leading chiefly to embryonic Hbs Gower-1 ($\zeta_2\varepsilon_2$), Gower-2 ($\alpha_2\varepsilon_2$), Portland-1 ($\zeta_2\gamma_2$) and Portland-2 ($\zeta_2\beta_2$). In mouse EryP cells, all *HBB*-like globins genes are expressed, initially (βH-1) (E 7.5), then on E 10.5 approximately equal levels of εy and βH-1 transcripts are detectable, until from E 15.5 the εy expression prevails as the cells terminally differentiate, at persistently minimal levels of adult β1- and β2- transcripts from E 10.5 [6,12].

1.4 Definitive erythropoiesis

Initially, definitive erythroid progenitors cells, erythroid burst forming units (BFU-E), emerge in the YS, from where they colonise the FL via the blood circulation (mouse E 9.5). BFU-Es are considered the most immature erythroid-restricted progenitors and expand in numbers in the FL, before they differentiate to more mature erythroid progenitor cells, erythroid colony forming units (CFU-E) (peak at mouse E 14.5–15.5) [6]. Both, BFU-E and CFU-E, are defined by their ability to form colonies when cultured in semi-solid methylcellulose supplemented with cytokines. Cultures of BFU-E colonies in methylcellulose give rise to thousands of haemoglobinised cells on day 5–8 (mouse) or 10–14 (human) in methylcellulose cultures. In steady-state conditions, BFU-E are found in the BM at frequencies of 40–120/10^5 BM cells and 10–40/10^5 light-density mononuclear cells in the peripheral blood [5,14]. CFU-E colonies are smaller, consisting of 16–125 cells, and appear on day 2–3 (mouse) and 5–8 (human) in semisolid cultures. CFU-Es are 5–8 fold more abundant in BM than BFU-E and are not found in circulation. As differentiation proceeds, foetal erythroid precursors associated with foetal macrophages form erythroblastic islands, where a gradual maturation of the precursors takes place.

Terminal erythroid differentiation begins when immature erythroblast differentiate to prepro-erythroblasts (PreproEr) further to Pro-erythroblast (ProEr). Maturation from ProEr to reticulocytes (Ret) is completed in three mitoses following the expected 1:2:4:8 ratio for ProEr to basophilic normoblast (Baso) to polychromatic normoblast (Poly) and finally orthochromatic normoblast (Ortho), respectively [15,16]. In every step of differentiation, there is a gradual reduction in cell size, karyopyknosis, RNA reduction and Hb accumulation. Eventually, enucleation takes place leading to the formation of the enucleated reticulocyte and pyrenocyte. Pyrenocytes hold the condensed nucleus in a thin rim of cytoplasm enclosed in a bilayer of cell membrane, exposing phosphatidylserine (PS) at the outer lipid layer. PS acts as an “eat me” flag for the macrophages [17] and thus triggers the phagocytosis of pyrenocytes, while reticulocytes enter the blood stream, where they finally develop into mature RBCs and remain for approximately 45 and 120 days for mouse and human, respectively. Reticulocyte maturation includes: reduction of plasma cell surface (~20%) and

cell volume, removal of residuals of internal organelles, including mitochondria and ribosomes, through autophagy and exocytosis, and reorganization of the cytoskeleton, leading to the formation of a mature biconcave RBC. Finally, senescent RBCs are recognised and destroyed by splenic macrophages [7].

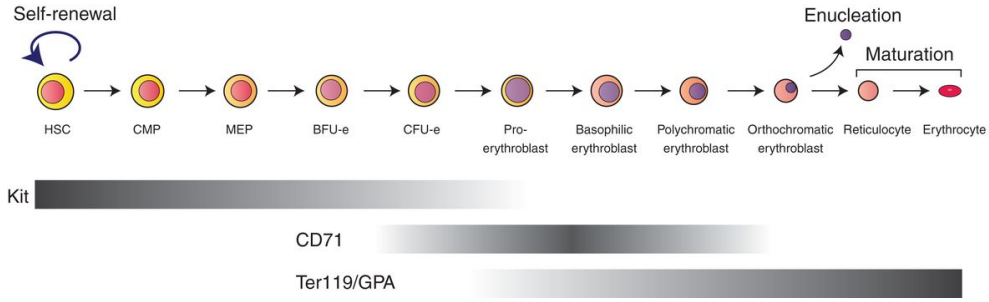


Figure 2: Erythroid differentiation.
Gradient bars designate the expression pattern of cell surface markers most commonly used for the identification of the different erythroid differentiation stages. Light grey, low expression; black, high expression); HSC, haematopoietic stem cell; CMP, common myeloid progenitor; MEP, megakaryocyte-erythroid progenitor; BFU-e, burst-forming unit, erythroid; CFU-e, colony-forming unit, erythroid. Reproduced from [5].

Definitive erythropoiesis in the FL initiates the production of foetal Hb (HbF), an $\alpha_2\gamma_2$ tetramer, as the number of EryD RBCs increases exponentially in the circulation in parallel with the rapid growth of the foetus [6]. The HbF tetramer allows the foetus to extract O_2 from maternal blood, because of its higher affinity for O_2 compared to adult Hbs [5]. This advantage of foetal over adult Hb is based on differences in their amino acid sequence (~72% identity) and particularly the substitution of a serine residue in foetal HBG- for histidine 143 in the adult HBB. This substitution reduces the binding affinity of HbF with 2, 3-bisphosphoglycerate (2, 3-BPG) compared to adult Hb. 2, 3-BPG is a highly anionic compound that is able to bind to the T form of deoxyhaemoglobin and stabilizes it, reducing effectively the binding of O_2. Therefore, the difference in O_2 binding affinity of HbF over HbA are essential for the effective transfer of O_2 from mother to foetus via the placental [18] thus providing a mechanism that favoured evolution of this so-called Hb switch in man.

Around the time of birth, the site of erythropoiesis switches from FL to BM, where HSCs expand and start a steady-state production and differentiation of definitive RBCs (Figure 3), before the second major transcriptional switch for HBB-like globin chains occurs in the EryD progenitors: the down-regulation of *HBG* expression and the initiation of adult *HBB* and *HBD* expression [5].

Briefly, adult haematopoiesis starts from HSCs located primarily in the aorta-gonad-mesonephros (AGM) region, from which they migrate to the FL, where they differentiate to mature HbF-expressing EryD before entering the blood. Postnatally, HbF ($\alpha_2\gamma_2$) is depleted to its adult steady-state level (of approximately 1% of total Hb), while adult Hbs HbA1 ($\alpha_2\beta_2$) and HbA2 ($\alpha_2\delta_2$) are up-regulated (to approximately 97% and 2% of total Hb, respectively) [19]. Initially, HSC clusters were found in close

association with endothelial cells in the ventral aspect of the aorta which led to the hypothesis that the permanent adult haematopoietic system is derived from "haemogenic" endothelial cells [5]. A recent study by Boisset *et al* confirmed this hypothesis by showing via live imaging of mid-gestation mouse aorta the real-time emergence of phenotypically defined HSCs (Sca1, c-kit, CD41) directly from the ventral aortic haemogenic endothelial cells [20]. Perinatally, HSCs migrate from the FL and colonize the BM, where they also make a transition from an actively cycling to a mainly quiescent cell state. The corresponding BM niche comprises endothelial cells of the vascular system, osteoblasts, stromal cells and the extracellular matrix. This microenvironment allows cell-to-cell interaction with the HSCs through adhesion molecules and their exposure to growth factors and cytokines essential for maintaining their multipotency and self-renewal capacitance, which are both equally important for long-term engraftment and haematopoietic reconstitution after BM transplantation or injury [5]. A subset of LT-HSCs divides to form a second HSCs population, the ST HSC, which have a more limited capacity for self-renewal but maintain their ability to give rise to all blood lineages of the adult organism. HSCs are gradually divided and differentiated to lineage-specific progenitors, including erythroid-specific progenitors, the adult-BFU-E and further CFU-E. Similar to foetal EryD erythropoiesis, EryD precursors in interaction with a central (nurse) macrophage form BM erythroblastic islands and progressively differentiate from proerythroblasts (ProEr) to basophilic (Baso), polychromatophilic (Poly) and orthochromatic (Ortho) normoblasts (Figure 2) [7]. Nurse macrophages are vital to the process of erythropoiesis, since they support the proliferation and survival of maturing EryD erythroblasts, help in the enucleation of Ortho and phagocytosis of the newly formed pyrenocytes, and are even implicated in iron recycling from senescent RBCs [17]. Reticulocytes are released from the erythroblastic islands into the blood stream, where they further mature into the biconcave RBC. The small size (6–8μm) and biconcave shape increase the cell surface for gas exchange and at the same time allow cells to travel in tissue microcapillaries (Figure 2) [5]. A schematic representation of the ontogeny of primitive and definitive erythropoiesis is shown in Figure 3.

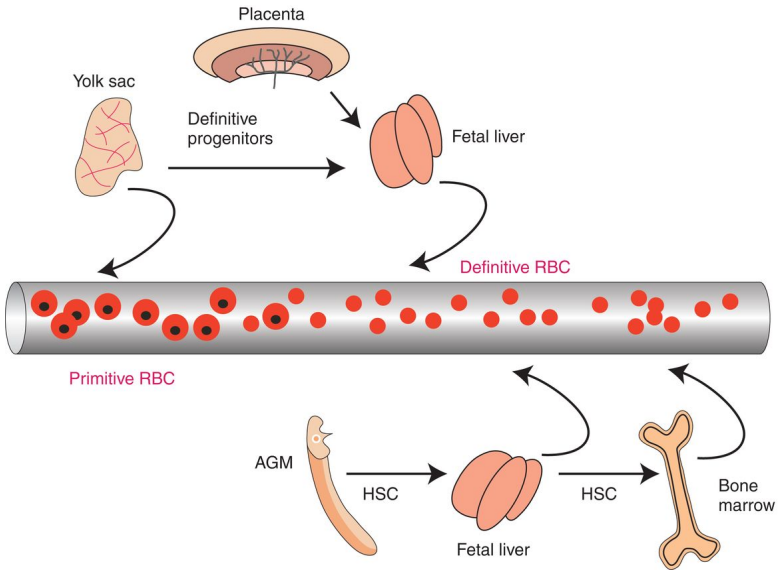


Figure 3: Ontogeny of erythropoiesis.
Primitive erythroblasts emerge in the yolk sac and enter the bloodstream as nucleated progenitors, which differentiate to mature primitive enucleated RBCs. Definitive erythroid progenitor cells (BFU-E) are generated in the yolk sac (YS) and placenta to then colonise the foetal liver (FL) via the blood circulation. In the FL, they differentiate to mature RBC and enter the circulation (expressing foetal/adult globin). The aorta-gonad-mesonephros (AGM) gives rise to the first HSCs, which migrate to the FL and enter circulation as mature definitive RBC. Foetal HSCs then colonised the bone marrow to provide life-long production of adult mature definitive RBCs in the circulation. AGM, aorta-gonad-mesonephros; HSC, haematopoietic stem cell; RBC, red blood cell. The horizontal cylinder symbolises blood circulating in blood vessels. For details, see main text. Figure reproduced from [5]

1.5 Haemoglobin

Haemoglobin (Hb) is the oxygen transporter of erythrocytes. Hb was first described by Friedrich Ludwig Hünefeld in 1840 as crystalline material in blood samples from earthworm held between two slides, and as plate-like crystals in dehydrated swine or human blood samples [21]. Later on, Max Perutz showed the three-dimensional structure of Hb using X-ray crystallography, a discovery for which he was awarded with the Chemistry Nobel prize in 1962 [22]. The structure is nearly spherical with a diameter of 5.5 nm; approximately $2.5x10^8$ molecules are packed per adult RBC [23,24]. Hb is a heterotetramer, normally consisting of two HBA-like globin chains and two HBB-like globin chains, with each subunit tightly associated with one haeme group. The main Hb function is the transfer of O_2 form lungs to tissues [21]. The tertiary structure of HBA and HBB chains includes seven and eight stretches of the α-helix, respectively, which when folded form compact globules and upon heterodimerisation form the Hb tetramer [23]. Each globin chain is bound to a prosthetic group, haeme, which is also responsible for the red colour of blood [18]. Haeme groups and the affinity of their central iron ion for one O_2 molecule, hence four per Hb, are essential for O_2 transport. Hb binds and transfers carbon monoxide (CO) and nitric oxide (NO) in a similar fashion, while CO_2 is instead transferred in blood in solution or by interacting with the amino-terminal residues of Hb as a weak carbamino complex [23].

Haeme is a protoporphyrin-IX ring carrying a central iron ion, which under normal conditions is in the ferrous oxidation state (Fe^{2+}) (Figure 4). The iron ion in haeme can make two additional bonds on the 5^{th} and 6^{th} coordination sites, of which the 5^{th} is occupied by the imidazole ring of a histidine residue in deoxyhaemoglobin molecule, leaving only the opposite 6^{th} site available for binding O_2. Iron in the ferrous state is relatively large and is placed slightly outside (~ 0.4 Å) of the protoporphyrin ring. With O_2 binding to the 6^{th} coordination site, a rearrangement of the electrons makes the iron slightly smaller, allowing its movement into the plane of the protoporphyrin ring. This alteration leads to the conformational changes of the structure of Hb moving from a tensed state (T) (deoxydative Hb) to relaxed state (R), when O_2 are bound to the Fe^{+2}, which leads to a ~15 degrees rotation of the $\alpha_1\beta_1$ dimer compared to the $\alpha_2\beta_2$ as seen in Figure 4 [18].

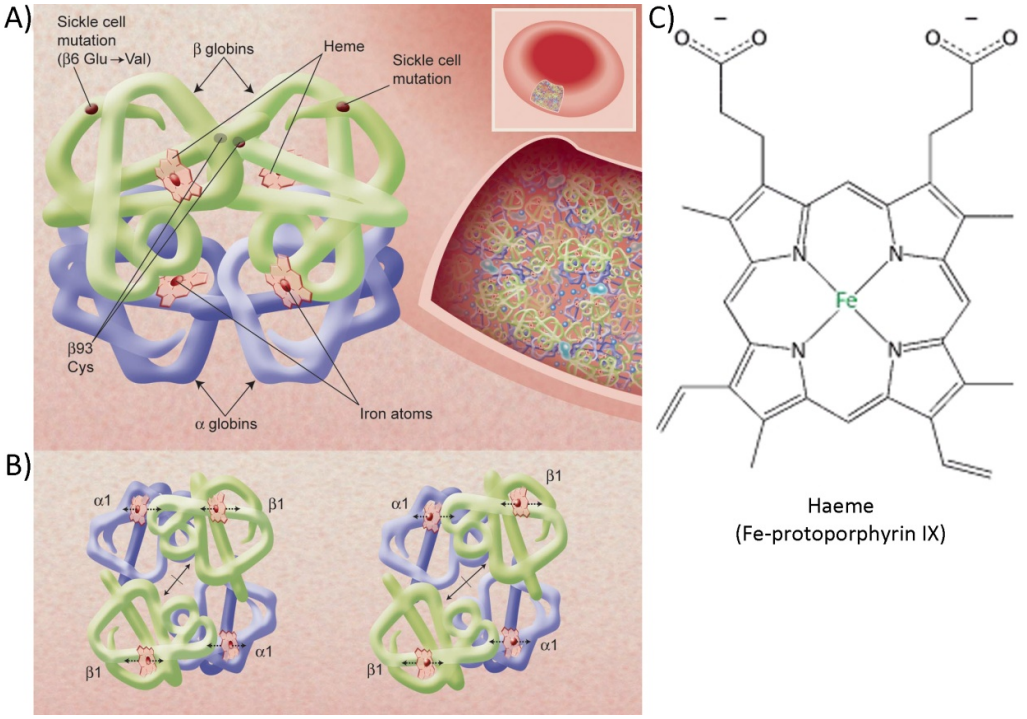


Figure 4: Haemoglobin structure.
A) Haemoglobin is a heterotetramer, normally consisting of two α-like globin chains (purple) and two β-like globin chains (green), with each subunit tightly associated with one haeme prosthetic group, a protoporphyrin-IX ring, (pink) carrying a central iron atom (red) responsible for gas binding. The sickle cell mutation (β6 Glu > Val) and the highly conserved β93 cysteine (NO binding) on the β-globin chains are shown. B) shows the conformational change of the structure of haemoglobin molecule between oxy-conformation (relaxed) and deoxy-conformation (tensed). Conformational change leads to a ~15 degrees rotation of the α1β1 dimer compared to the α2β2. (A) and (B) modified from [23]. C) Structure of Haeme Fe-protoporphyrin IX ring with the central Fe atom designated in green. Image modified from [18].

Hb is a great example of an allosteric protein and exhibits a cooperative binding of oxygen to Fe^{+2}. In particular, the binding of O_2 to one site leads to conformational changes of the protein, which increases the binding affinity of the neighbouring globin chains for O_2 (approximately threefold relative to the binding affinity in fully deoxygenated deoxyhaemoglobin). Cooperative binding of O_2 by Hb allows the delivery of the 1.7-fold amount of O_2 to tissue compared to (hypothetical)

independent O_2 binding. In addition, decreased pH and increased concentration of CO_2, conditions commonly found in muscle tissues, lower binding affinity for O_2 and actually promote its release. In general, the transition of Hb between tensed and relaxed state is regulated based on the pH and the O_2 and CO_2 pressure of the surrounding environment [18].

1.6 Globin expression

Hb expression is controlled by two globin-gene clusters, an *HBA*-like globin cluster , which is located at the short arm of chromosome 16 close (~150 kb) to the telomere, in an early replicating GC-rich region, with ubiquitously active genes, and a *HBB*-like globin cluster, located on the short arm of chromosome 11 in a late replicating region, methylated with a relatively low number of genes, which in contrast to the *HBA* cluster chromatin environment is maintained as closed-chromatin (heterochromatin) in non-erythroid cells [25,26] Difference in the chromatin environment between the two globin clusters in non-erythroid cells can be explained by the *HBA* cluster being surrounded by housekeeping genes (e.g. nitrogen permease regulator 3-like protein (*NPRL3*)), [27] which is transcriptionally active, regardless of the lineage specificity [25,28].

The *HBA*-like cluster in all mammalian species contains a ζ-globin gene (HBZ, embryonic globin) followed by a duplicated α-globin gene for adult globins α1 and α2 (HBA1 and HBA2, respectively) (Figure 5). In addition, the human locus includes two pseudogenes, αD and θ, with the former proceeding and the latter following the adult *HBA2* and *HBA1* (5' – ζ – αD – α2 – α1 – θ – 3') [27]. In general, in all the different developmental stages the expression of HBA-like and HBB-like chains is balanced [29]. The coordinated expression of the globin genes at each developmental stage is dependent on the regulatory elements located in distant regions upstream from each cluster and characterised by hypersensitivity to DNase I nuclease in erythroid cells. In general it have been shown that DNase I hypersensitive regions in many mammalian genes are associated with sites that are involved in protein-DNA interactions, such as enhancers and promoters [25,30]. The *HBA* cluster holds four regulatory elements (enhancers), the multispecies conserved sequence (MCS) R4 (MCS-R4) (10 kb), the MCS-R3 (33 kb upstream), the MCS-R2 (40 kb upstream) and the MCS-R1 (48 kb upstream from the α-like-globin genes), with the MCS-R2, aka HS40 playing the most prominent role in regulation of HBA-like chain expression (Figure 5) [25,26]. In general, both *HBA* genes are continuously expressed from erythroid cells from early stages of embryonic erythropoiesis. In addition, *HBZ*, which is at 5' end of the *HBA* genes, is transiently expressed prior to *HBA* at the early embryonic stages (until the 6th week of gestation).

The *HBB*-like cluster consists of five *HBB*-like genes, the embryonic ε (*HBE*), the foetal ^{G}γ (*HBG2*) and ^{A}γ (*HBG1*), and the adult δ (*HBD*) and β (*HBB*) genes, arranged on chromosome 11 relative to their own regulatory region, the *HBB* locus control region (βLCR), in the order of their developmental expression [19,29,31] (Figure 5). Finally, between *HBG1* and *HBD* there is also ψβ pseudogene that is not expressed (5' – *HBE* – *HBG2* – *HBG1* – ψβ – *HBD* – *HBB* – 3') (Figure 5). The βLCR is located 6 to 22 kb upstream of *HBE* and comprises of five DNase I hypersensitive site (HS) elements (5' HS5 - HS1 3'), which are essential for the normal high-level expression of the HBB-like chains and for their developmental regulation [32]. A 3' HS region was also identified, known as 3'HS1, which seems to play a role in the formation of the active chromatin hub (ACH), but elicitation of its precise role requires further study [19]. HS1–4 are erythroid-specific and act as enhancer-like elements, whereas the HS5 acts as non-specific insulator [33]. HS core elements are 200–400 bp in size, with more than 2 kb separating each HS from the next, and contain highly conserved binding sequences of several ubiquitous and lineage-specific transcription factors, such as MARE (Maf recognition element) and (A/T GATA A/G) GATA sequences found in HS2, 3 and 4, CCAAT-containing likely KLF binding sites in HS2 and 3, and an E-box motif in HS2, in which USF and Tal 1 factors and its heteromeric partners seem to interact, all known to play a key role in erythropoiesis [32,34]. Besides DNA-protein interaction, protein-protein interactions also play a significant role in βLCR function, such as those involving co-activators and acetyltransferases, in part required for recruitment of macromolecular complexes that perform chromatin remodelling and histone tail modifications [32]. HS2 is a classical strong enhancer element that is equally active in all developmental stages for all *HBB*-like genes. Both transient and stable transfection assays indicated HS2 activity. This contrasts with the HS3 and HS4 elements, which were shown to only exert transcriptional enhancement activity from within a stably integrated chromatin context [33]. The βLCR is a highly potent enhancer and regulator of the *HBB*-like genes, with the ability to overcome heterochromatin suppression and to provide an open chromatin structure and dominant transcriptional activating function for the whole *HBB* locus in erythroid cells. The βLCR is required for high expression from the locus during erythroid development and is also critical to differential regulation of expression of different *HBB*-like genes, which involves long-distance interaction with globin promoters by chromatin looping with the assistance of looping-protein complexes. This can be demonstrated, for example, by the βLCR providing position-independent and copy-number-dependent expression of linked globin transgenes in transgenic mice. At the 5'end of each *HBB*-like gene a specific promoter and corresponding *cis*-acting regulatory elements such as ATA, CCAAT, and CACCC sequences, regulate the binding of a complex of proteins that control the initiation and rate of transcription [23]. In addition a (A/T GATA A/G) GATA binding

sequence is found in multiple sites throughout the *HBB*-like cluster. GATA-1 is an erythroid-specific transcription factor with a major role in primitive and definitive erythropoiesis [7].

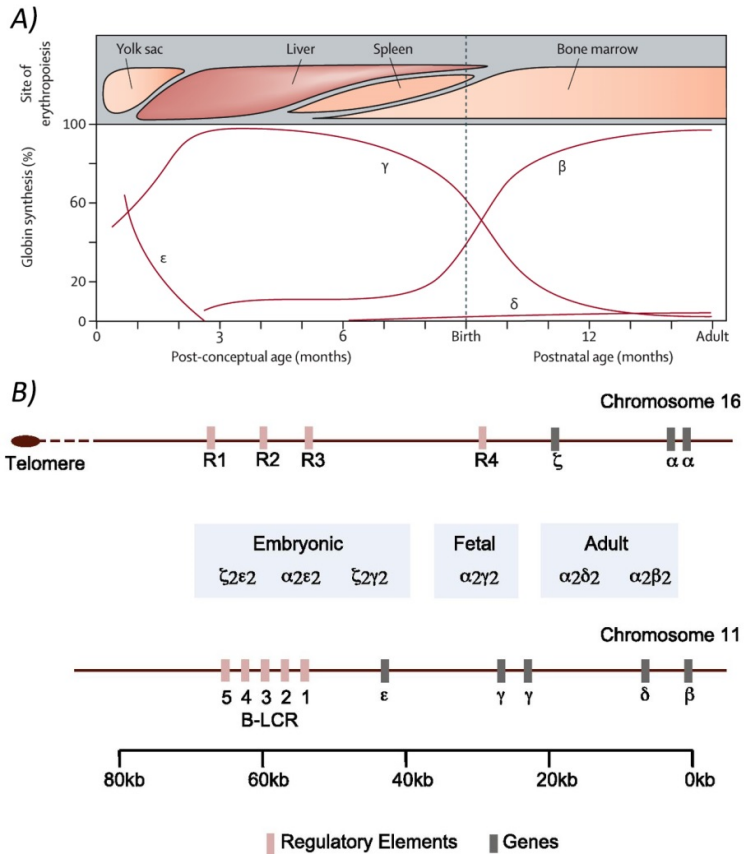


Figure 5: Regulation of globin gene expression.
A) Top: Sites of erythropoiesis and corresponding expression of HBB-like chains during development. Figure reproduced from [29]. B) Structures of *HBA*-like and *HBB*-like gene clusters on chromosomes 16 and 11, respectively. The *HBA*-like cluster is located at the short arm of chromosome 16 close to the telomere and holds four regulatory elements, the multispecies conserved sequence (MCS) R4 (MCS-R4) (10 kb), the MCS-R3 (33 kb upstream), the MCS-R2 (40 kb upstream) and the MCS-R1 (48 kb upstream from the *HBA* genes) and three functional *HBA*-like genes (ζ, *HBZ*; embryonic) and duplication of *α* (*HBA1* and *HBA2*; foetal/adult)). The *HBB*-like cluster is located on chromosome 11 and is associated with five regulatory elements, which together comprise the HBB-locus control region (βLCR) and five functional *HBB*-like genes (ε, *HBE*; embryonic), two γ (*HBG1* and *HBG2*; foetal), δ (*HBD*; adult) and β (*HBB*; adult). Globin genes are arranged in a manner corresponding to the pattern of their developmental-specific expression, leading to the production of embryonic ($\zeta_2\varepsilon_2$, $\alpha_2\varepsilon_2$ and $\zeta_2\gamma_2$), foetal ($\alpha_2\gamma_2$) and adult ($\alpha_2\delta_2$ and $\alpha_2\beta_2$) Hbs. Figure reproduced from [26].

1.6.1 *HBB-like gene switching*

As detailed above in relation to development and anatomical structures during ontogenesis, two HBB switches occur in human development. Initially during primitive erythropoiesis in the YS, only *HBE* is expressed until the 6-8th week of gestation, when erythropoiesis is transferred to the FL. There, the first switch takes place, during which *HBG1* and *HBG2* are upregulated and *HBE* is silenced. Foetal HBG chains from the two genes are functionally identical, and as a result of gene duplication differ only in amino acid position 136, which holds a glycine in HBG2 and an alanine in HBG1 [35]. The second switch occurs around the time of birth, when the foetal globins are

downregulated and adult *HBD* and *HBB* are activated for continued expression from BM and spleen throughout postnatal life. Of note, *HBD* misses essential control elements within its promoter for high-level expression and is therefore only expressed at low levels throughout adult life [36,37]. Switching from foetal Hb to adult is completed during the first months of life, leading to a steady-state level of 97% HbA, 2% HbA_2 and 1% HbF [23] at the end of a development that through overlapping developmental expression patterns for different globin genes has seen the formation of a variety of structurally defined Hbs [38]. At that stage, residual expression of HbF is normally 1%; a product of a subset of erythrocytes that continue to express the *HBG* genes and are known as F-cells (see Figure 5) [5,19].

Molecular control of the HBB-like switch has long served as a model for eukaryotic gene regulation during development. Additionally, high-level HBG expression is partially or fully therapeutic in β-globinopathies, so that over the last two decades fundamental research has been complemented with translational research into a potential therapeutic exploitation of the switch from foetal to adult globin expression. As a result, an exceptional level of characterisation of the molecular mechanisms of Hb switching has been attained, and a still growing array of transcription factors, miRNAs and cis-acting elements have been unearthed that participate in its control [39]. HS upstream regulatory elements are accessible by transcription factors, activators or repressors, depending on the developmental stage of the organism, where they specifically activate or repress the expression of the globin genes. This can occur by altering the methylation pattern of DNA, where the presence of the methyl groups of 5m-cytosine in the major grove of DNA interfere with binding of transcription factors, so that methylation is associated with inactivation of gene expression. In addition, methylation may recruit additionally inactivating histone-modifying factors, such as histone deacetylases, with further epigenetic effects on gene expression. Generally, methylation and deacetylation of particular histone residues can change the formation of the chromatin to a closed-structure chromatin, heterochromatin, where transcription factors do not have access to their DNA binding sequence, therefore repressing expression. Acetylation of some other histone residues in combination with other epigenetic changes may have the opposite effect and lead to an opened-chromatin structure, euchromatin, which is linked with high expression of the adjacent genes [29]. In each instance the combined effect of epigenetic modifications has to be analysed for the locus in question. By default and barring epigenetic modifications of the *HBB* locus, the promoter most proximal to the βLCR is activated, so that developmental switching from βLCR-proximal (upstream) to βLCR-distal (downstream) promoters requires active regulation and specific regulatory elements in association with each promoter, to modulate interaction with transcription factors and regulatory βLCR elements [29,40,41]. The βLCR HS elements act upon the *HBB*-like promoters, amongst other

mechanisms, indirectly by binding auxiliary transcription factors or directly by the binding of looping protein complexes and direct interaction. Of particular importance are HS1 through 4, which each recruit specific transcription factors but *in toto* contribute in an additive fashion to recruit the *HBB* locus to RNA pol2-associated transcription factories in the central nucleoplasm [42]. Several studies have shown evidence for the presence of a competition for βLCR elements between *HBB*-like globin genes. For instance, deletions of the *HBB* promoter elements, CACCC, CCAAT and TATA boxes, in heterozygotes were associated with increased levels of HbA2 ($\alpha_2\delta_2$) and variable levels of HbF ($\alpha_2\gamma_2$) [43]. Similar observations have been made in carriers for point mutations that inactivate the *HBB* promoter [40]. In the case of homozygosity for deletions of the *HBB* promoter, two individuals were transfusion-independent with mild β-thalassaemia, HbF being significantly increased and compensating for the absence of HBB chains [41,44,45]. Therefore, inactivation of the *HBB* promoter shifts the competition between *HBB*-like genes for transcription factors and for the βLCR towards activation of alternative HBB-like promoters. Where downstream promoters are mutated, this can therefore partially revert developmental globin switching back to upstream promoters [44,45].

1.7 Globin disorders

In a wider sense, haemoglobinopathies are any gene defects that lead to lowered levels of functional Hb, be it by structural or regulatory defects, and taken together have a global carrier frequency of 7% [46,47]. Globin disorders in general are prevalent in tropical and subtropical regions where malaria has been or still is endemic, owing to the natural protection of heterozygotes against the parasite *Plasmodium falciparum* [25,29,48]. However, due to global migration these disorders are becoming more common in countries today where they have previously been rare, such as in northern Europe, Australia and North America [29].

While many use the terms 'haemoglobinopathy' and 'globin disorder' interchangeably, some authors distinguish between thalassaemias, as quantitative disorders where Hb production is reduced or abolished, and haemoglobinopathies as qualitative disorders resulting from defects of globin structure, which may in turn lead to the formation of abnormal Hb [46]. Following is a brief introduction to haemoglobinopathies in the strict sense of structural abnormalities, before briefly turning to α-thalassaemia and expanding on β-thalassaemia as the main focus of this thesis.

1.7.1 *Structural abnormalities*

More than 700 variants of Hbs have been identified, of which structural haemoglobinopathies are inherited as autosomal co-dominant traits. Of particular frequency and clinical importance amongst

this class of mutations are those giving rise to sickling Hb (HbS), causing the severe sickle cell disease (SCD), and the far milder HbE and HbC [49].

SCD, for instance, is characterized by the production of an aberrant HBB protein (β^S) caused by a codon 6 (GAG>GTG) mutation (human genome variation society (HGVS): HBB: c20A>T), which leads to an amino acid substitution ($HBB^{Glu6Val}$). The resulting HbS haemoglobin tetramer polymerizes when de-oxygenated, with severe health consequences [49]. Left untreated, patients with this globally most common haemoglobinopathy die before the age of 5 years [50]. Another frequent variant, HbC, is also produced by a codon 6 (GAG>AAG) mutation (HGVS: HBB:c.19G>A), which leads to an alternative codon substitution (HBBGlu6Lys) and to less severe disease progression than SCD, causing mild haemolytic anaemia [47,51]. HbE on the other hand, is characterised by a reduced production of a structurally aberrant β^E-globin protein, caused by a codon 26 mutation (GAG>AAG) (HGVS: HBB: c.79G>A) and corresponding activation of cryptic splice site, which leads to reduction of β^E-globin synthesis. Co-inheritance of structural globin abnormalities, such as HbE, HbC and HbS, with the thalassaemias lead to complex clinical phenotypes and are of great regional importance, as may be exemplified by β-thalassaemia/HbE in Southeast Asia [47].

1.7.2 *α-Thalassaemia*

A comprehensive review of α-thalassaemia is beyond the scope of this introduction and has been performed elsewhere [52]. In brief, approximately 5% of the world population are carriers for α-thalassaemia mutations, which reduce or completely abolished the expression of HBA chains [53]. Based on the IthaGenes database, approximately 258 mutations for α-thalassaemia have been identified, most of which are deletions [54]. Normal genotype is αα/αα with genes *HBA1* and *HBA2* encoding two identical HBA chains as indicated in Figure 5. α-Thalassaemia mutations are categorised into two subgroups: α^+, in which usually one of the linked *HBA* genes is affected (α-/αα for heterozygotes) and α^0 in which both are affected (--/αα for heterozygotes). Individuals with one *HBA* affected are clinically silent, asymptomatic, and might present mild hypochromia in RBCs and marginal reduction in Hb values. Conversely, in minor α-thalassaemia, also known as α-thalassaemia trait, two *HBA* genes are affected, which might be the case for homozygotes of α^+ thalassaemia (-α/-α) or heterozygote of α^0 thalassaemia (- - / αα), leading to mild, hypochromatic microcytic anaemia [47,53]. In compound heterozygotes of α^+/α^0 thalassaemia, where usually three out of four *HBA* genes are inactive (α-/--), the HBA-to-HBB chain balance is changed significantly, so that we have the formation of extremely unstable β_4 tetramers (HbH) with high O_2 affinity. This brings about inefficient O_2 transfer and haemolytic anaemia with ineffective erythropoiesis [47], so that HbH disease in general causes moderate hypochromic haemolytic anaemia with splenomegaly. Finally, in the case

of homozygous α^0 thalassaemia (--/--), there is no production of HBA chains, and with the gradual increase of HBG *in utero*, free HBG chains accumulate to form γ_4 tetramers, also known as Hb Bart's. Hb Bart's has high O_2 affinity and is relatively insoluble and stable, the resulting low oxygen delivery and accumulation in RBCs leading to *hydrops fetalis*, which leads to pre-natal death of the foetus [47].

1.7.3 *β-Thalassaemia*

1.7.3.1 *Epidemiology and genetics*

β-Thalassaemia is caused by mutations that lead to the reduction or complete abolishment of expression of functional HBB chains. Over 80 million people (1.5%) worldwide may be carriers for β thalassaemia, and it is estimated that 60 000 affected infants are born every year [29,55]. Over 400 mutations causative of β-thalassaemia are currently held on the ITHANET portal (http://www.ithanet.eu/db/ithagenes) [54]. Globally, Cyprus and Sardinia have the highest frequencies of β-thalassaemia carriers, with 14% (12% according to updated information on the ITHANET portal) and 10.3% of the population, respectively [28].

Complete absence of HBB is designated $\beta^{0,}$ whereas reduction of functional HBB is designated as β^+ or, if negligible, β^{++} [56]. Homozygotes and compound heterozygotes for the above alleles result in a wide spectrum of thalassaemic phenotypes which largely correspond to the clinically defined phenotypes major ($\beta^{0/0}$), intermedia (β^0/β^+ or β^+/β^+) and minor (β^{++}/β^{++}) [56,57].

Most β-thalassaemia mutations are non-deletional, including single-base substitutions or small insertions and deletions within *HBB* or the immediate flanking sequence, and have been shown to be involved in all stages of *HBB* expression, from transcription over to RNA processing to translation [56]. In general, point mutations in the 5' untranslated region (UTR) and the promoter lead to defective *HBB* transcription. Mutations affecting RNA processing include i) mutations altering the GT and AG invariant splice sites at the intron borders, which have mostly been designated as β^0, ii) mutations of relatively conserved consensus sequences near border regions of introns, which lower the efficiency of normal splicing similar to iii) mutations activating cryptic splice sites or iv) mutations creating abnormal splice sites (mostly β^+), and v) mutations affecting the poly A signal and 3'UTR with relatively minor impact on *HBB* expression (mostly β^{++}). When translation is affected either by disruption of the initiation codon or by nonsense mutations and frame-shifts, the mutations are usually designated as β^0. Rare large deletions have also been identified as responsible for β-thalassaemia and may be divided into those that affect the βLCR together with *HBB*-like globin genes and those that affect the βLCR alone.

In rare cases, *trans*-acting mutations have also been shown to cause β-thalassaemia, such as mutations on the XPD subunit of the TFIIH transcription factor or of the erythroid-specific transcription factor GATA-1 [56,58,59]. Most common β-thalassaemia mutations are recessive and cause haploinsufficiency, while rare cases of dominant inheritance involve highly unstable variant proteins that form toxic insoluble precipitates in the erythroid precursors even in the heterozygous state. These mutations combine aspects of qualitative and quantitative Hb disorders and in most cases lead to the formation of a pre-termination codon in the 3' end of exons 2 or in exon 3, so that relatively non-functional, unstable HBB chain variants are produced that escape the surveillance mechanism of the non-sense mediated decay (NMD) pathway (see 1.8). RBCs of patients with recessive β-thalassaemia mutations contain inclusion bodies that are mostly composed of precipitated excess HBA chains, while for dominant mutations the unstable HBB chains contribute to inclusion-body formation and thus lead to increased ineffective erythropoiesis [56,60].

1.7.3.2 *Pathophysiology*

Thalassaemia major causes severe anaemia, which becomes apparent 3–6 months after birth, when the switch from foetal globin to adult is almost completed and the predominant Hbs are HbA and HbA2 [19]. Thalassaemia-major patients require regular blood transfusions early-on, which in turn cause organ iron overload, particularly in heart, liver and other endocrine glands, impairing their function. In the absence of blood transfusions, patients die within the first two decades of life, while complete compliance to a costly and often invasive regimen of blood transfusions and iron-chelation therapy allows even pregnancies and survival beyond the age of 40 [57]. As major consequences of ineffective erythropoiesis and iron overload, patients with β-thalassaemia present hepatosplenomegaly owing to increased destruction of RBCs and with extramedullary haematopoiesis (EMH) and bone deformities owing to BM expansion [57,61]. The latter is caused by increased numbers of erythroid precursors and accumulation of countless apoptotic cells at the polychromatic and orthochromatic stages, at approximately 15 times the levels found in healthy individual [62]. Iron overload constitutes a major and unavoidable complication of β thalassaemia, caused in part by lifelong regular blood transfusions and in part by increased gastrointestinal iron absorption secondary to inhibition of hepsidin synthesis [63]. Consequently, 71% of deaths for thalassaemia major are due to cardiac diseases, rendering iron chelation treatment mandatory [64]. Approximately 6.8% and 5.7% of the thalassaemia major patients present heart failure and arrhythmias respectively. Furthermore, 70% of patients developed liver diseases due to viral infections with chronic hepatitis B and/or C, and in association with liver iron overload post-transfusion, there is an increase of the risk to developed cirrhosis and hepatocarcinoma [65].

Furthermore, venous thrombosis, hypogonadism, diabetes and osteoporosis constitute some of the common complications that can be developed in β-thalassaemia patients [57]. Thalassaemia intermedia covers variable degrees of severity, ranging from severe to mild anaemia, and does not (or only occasionally) require blood transfusion. In general, severe forms of β-thalassaemia intermedia are clinically manifested between the ages of 2–6 years while for mild forms, patients can be completely asymptomatic into adulthood. While transfusions in thalassaemia major partially suppress splenomegaly and EMH and consequential bone deformities, with exception of those caused by iron overload, these disease features are common for severe forms of thalassaemia intermedia [57]. Finally, β-thalassaemia minor is usually asymptomatic or presents with mild microcytic hypochromic anaemias with increased HbA2 [47,57].

1.7.3.3 *Modifiers*

The main parameter determining the severity of thalassaemia is the severity of the underlying primary mutation(s) (see above). However, between different patients with the same primary mutation, there is great phenotypic variability of disease severity, a phenomenon brought about by disease modifiers. The molecular mechanism underlying thalassaemia pathophysiology is a) anaemia and b) toxicity brought about by the relative excess of the unaffected globin chain. Disease severity is therefore a function of functional Hb levels for anaemia and of the relative imbalance of HBA and HBB chains for haemolysis, making the abundance of the respective binding partner in HbA the most important disease modifier for the latter. For instance, co-inheritance of HBA duplications increases the severity of β-thalassaemia, while co-inheritance of α-thalassaemia reduces it [61,65]. In β-thalassaemia, excess HBA chains impair erythroid maturation and subsequently precipitate in mature erythrocytes to cause haemolysis [29]. These primary changes lead to the clinical symptoms discussed above, which cause organ failure long-term [29,61]. The second category of major modifiers for β-thalassaemia is mutations that modify the imbalance of HBA and HBB-like chains by affecting *HBG* expression. Like co-inheritance of α-thalassaemia, increased expression of *HBG* reduces the globin-chain imbalance in β-thalassaemia, in this case by the formation of foetal Hb HbF ($\alpha_2\gamma_2$) [66]. Corresponding modifiers may be subdivided into mutations on the *HBB*-locus itself, which are traditionally referred to as hereditary persistence of foetal Hb (HPFH), and mutations affecting *trans*-acting factors, such as the transcription factors BCL11A, KLF1 (also known as EKLF) and SOX6, whose deficiency is associated with elevated *HBG* expression [61,67–69]. Finally, tertiary modifiers are loci that are not associated with globin expression, but that contribute to the complications caused by thalassaemia, such as iron overload (Human haemochromatosis protein HFE mutations HFE^{C282Y} and HFE^{H63D} mutation), gallstones (TA_7 polymorphic variant in the promoter of the uridine diphosphate-

glucuronyltransferase IA (*UGTIA*) gene), osteoporosis (mutations in the oestrogen receptor gene, vitamin D receptor (*VDR*), collagen type a1 (*COL1A1*) and *COL1A2* genes, and transforming growth factor b1 (*TGFb1*) and cardiac complications (decreased activity of apolipoproteins E4) [61].

1.7.3.4 *Cellular pathophysiology*

Under normal circumstances, free intracellular HBA chains are stabilized by the erythroid-specific molecular chaperon α-haemoglobin-stabilizing protein (AHSP), which promotes protein folding and resistance to protease digestion [70,71]. In the pathological condition of β-thalassaemia, however, the excess of free HBA chains saturates the stabilizing system, resulting in auto-oxidisation of the unstable free HBA chains and the formation of α-haemichromes and reactive oxygen species (ROS). The α-haemichromes are degraded to haemin and free iron, which precipitates on cell membranes through three different pathways (the adenosine-triphosphate- and ubiquitin-dependent proteolytic pathway, an autophagy pathway and a non-enzymatic pathway triggered by ROS). Haemin and free iron lead to the formation of more ROS through redox reactions that damage cellular proteins, lipids and proteins. Furthermore, oxidization of the band 4.1 protein leads to the impaired assembly of the membrane cytoskeleton complex (spectrin/actin/band 4.1) which is believed to be the main cause of RBC haemolysis. In addition, clusters of α-haemichrome bound by band 3 are recognised by naturally occurring anti-band 3 IgG and stimulate the alternative complement pathway for removal of oxidatively stressed RBCs from blood circulation by macrophages [72] . Ineffective erythropoiesis in thalassaemia can be explained as result of increased apoptosis, which is triggered through the FAS/FAS receptor pathway in response to increased levels of ROS. Moreover, the heat-shock protein 70 (hsp70) interacts with free HBA chains and becomes sequestered in the cytoplasm, leaving GATA 1 unprotected and permissive to proteolytic degradations, which in turn halts erythroid differentiation and promotes apoptosis of Poly (Figure 6) [26].

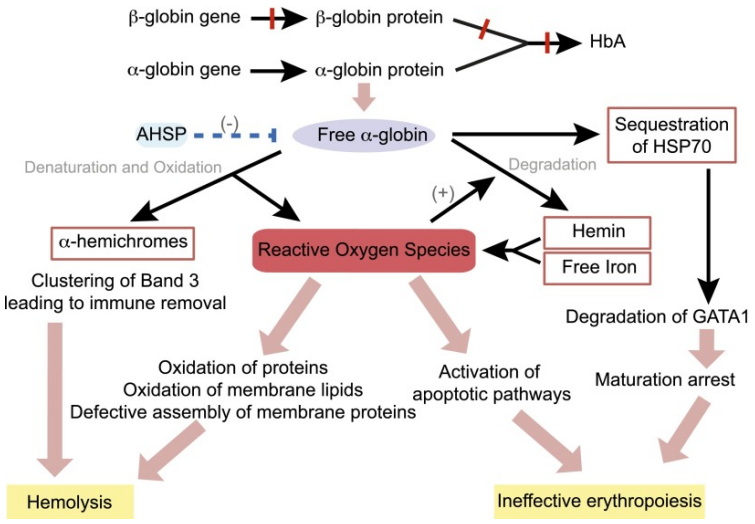


Figure 6: Cellular pathophysiology of β-thalassaemia.
Reduction of β-globin chain synthesis in β-thalassaemia, leads to imbalance of the α-/β-like ratio and reduction of the formation of functional haemoglobin A (HbA). The excess free α-globin chains trigger a cascade of events through the generation of ROS, which eventually causes the haemolysis of red blood cells and ineffective erythropoiesis. AHSP, α-haemoglobin stabilizing protein; HSP70, heat shock protein 70. Figure reproduced from [26]

1.8 Defects in RNA processing as a cause of disease

1.8.1 *RNA processing and human disease*

RNA processing is tightly associated with transcription, in that phosphorylation of the C' terminal domain of RNA polymerase (RNA pol) 2, in particular of Ser 5 and Ser 2 at the initiation and elongation steps, respectively, is as important for transcription as for the RNA processing [73]. Initially, the first processing event for each mRNA is the 5' end capping which occurs co-transcriptionally in the nucleus, as soon as the first 25–30 nucleotides of the pre-mRNA are transcribed [73]. Capping occurs in three enzymatic reactions, first, an RNA triphosphatase removes the γ-phosphate from the terminal nucleotide; second, a guanylytransferase uses GTP to form the GpppN, which is subsequently methylated by a guanine-N7 methyltransferase, at position N7 to form the Gmet7 end, also referred to as cap 0. Third and finally, methylation of the +1 and +2 ribonucleotide at the 2' O of ribose generates the cap +1 and +2, respectively [74]. Gme7 capping is essential for the stability of the mRNA from 5'>3' exonucleases, for efficient translation via recruitment of initiation translational factors (eiF4F complex) and for splicing, by recruitment of cap-binding complexes which are involved in spliceosome assembly, 3' end processing, RNA export and nonsense-mediated decay [74]. Cap +1 acts as a self-signature, for recognition and discrimination of mRNA from foreign (e.g. viral) transcripts, the absence of which triggers a protective mechanism to hold RNA translation [74].

Another co-transcriptional process of mRNA maturation is splicing, the removal of intervening (intronic) from exonic sequences by a ribonucleoprotein complex named spliceosome. There are four main elements in the pre-mRNA sequence which regulate splicing, the 5' [GU] donor splice (SD) site, the [AG] 3' acceptor splice (SA) site, the [A] branch point site (BPS) (10–30 nt upstream from the SA) and the poly-pyrimidine tract (PYT) (between the BPS and SA) (Figure 7) [75]. Splicing is completed with two transesterification reactions as follows. Initially the 2' hydroxyl group of the BPS adenosine residue makes a nucleophilic attack on the SD phosphate group, leading to the formation of a lariat (loop) intermediate, with 2',5'-phosphodiester linkage, followed by a second attack on the SA phosphate group by the 3'-hydroxyl group of the cleaved 5' exon, leading to the release of the intron lariat and ligation of the two exons (Figure 7) [75]. Identification and regulation of splicing is accomplished by the spliceosome complex, which contains five small-nuclear ribonucleoproteins (snRNPs), specifically U1, U2, U4, U5, U6 in the case of the major spliceosome (responsible for the splicing of ~95.5% of all introns) and U11, U12, $U4_{atac,}$U5 and $U6_{atac}$ for the minor spliceosome, which is present in only a subset of eukaryotes [76] . Initially, U1 snRNP is recruited to the SD and the non-snRNP factors SF1/BBP and U2 auxiliary factor (U2AF) bind cooperatively to the BPS and PYT, respectively, forming the E complex. SF1/BBP protein interacts with U2AF 65kDa subunit of U2AF,

which binds the PYT and indirectly with the U2AF 35kDa subunit, which in turn binds the SA AG di-nucleotide [77]. Subsequent to the formation of the E complex, the U2 snRNP stably binds the BPS, and SF1 factor is released, leading to the formation of the A complex, also known as pre-spliceosome. Subsequently, the U4/U6 snRNP is associated with U5 to form the U4/6.U5 tri-snRNP, and recruited to form the pre-catalytic complex B. Conformation changes and rearrangements lead to the release of U1 and U4 snRNP, which constitute the catalytically active B complex (B* complex), ready to complete the 1st step of splicing with the assistance of the DEAH-box RNA helicase Prp2. Then, the C complex is formed and further rearrangements lead to the 2nd catalytic step of splicing. Finally the spliceosome dissociates and the snRNPs are released to take part in additional rounds of splicing (Figure 8)[76].

Except from the main splice sites and BPSs, splicing is regulated by an array of other cis-acting regulatory elements, exonic and intronic enhancers (ESEs and ISEs) and silencers (ESSs and ISSs), which are the binding sequences of trans-acting regulatory splicing factors such serine/arginine (SR)-rich protein and heterogeneous nuclear ribonucleoproteins (hrRNPs). In general, SR proteins act as enhancers of splicing, whereas hrRNPs act as silencers, and both influence the recruitment and assembly of the spliceosome at particular splice-sites and in the definition of exon and intron (Figure 7)[76–79].

The last two steps of nuclear mRNA processing before nuclear export represent 3'-end maturation. The first step is endonucleolytic cleavage 10–30 nucleotides downstream of the conserved mammalian poly A signal, a usually AAUAAA adenylate-rich hexameric element, and upstream of the U/GU-rich sequence, the second is poly-adenylation of the cleaved-end by poly (A) polymerase, which is part of the complex that recognises the poly A core elements and also performs the RNA cleavage [73,80]. Poly(A) tailing is essential for RNA stability, RNA transcription termination, mRNA export from nucleus to cytoplasm and translation [73].

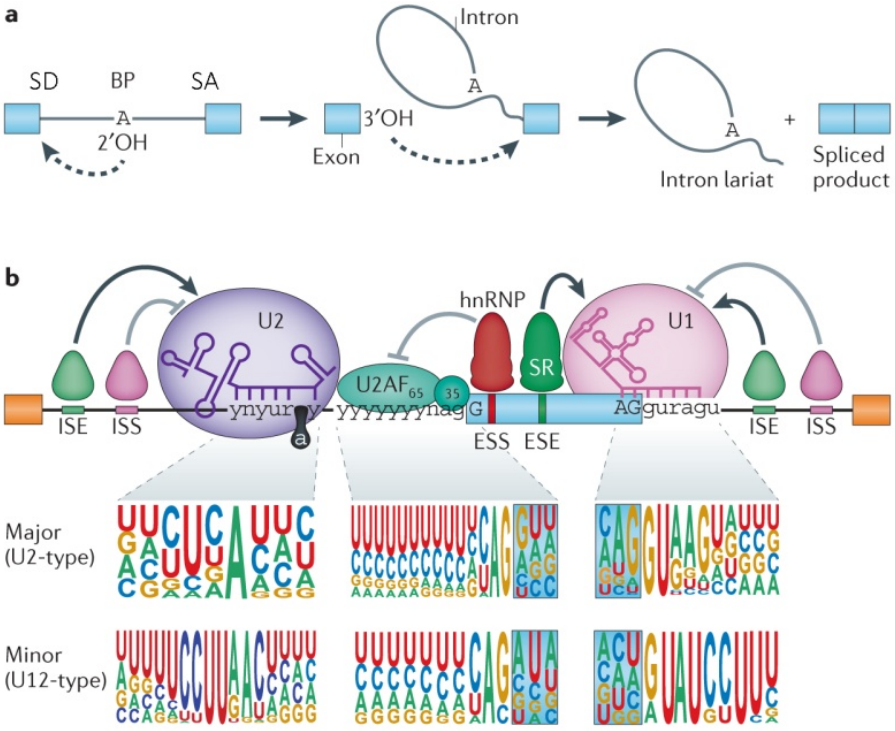


Figure 7: RNA splicing regulation.

a) Schematic representation of the two-step transesterification splicing reactions, initially with the 1st nucleophilic attack of the 2′-OH group of the adenosine residue of the branch point site (BP) on the phosphate group of the 5′ splice site (SD) and subsequently with the 2nd transesterification reaction, in which the 3′-OH group of the 3′ end exon attacks the SA phosphate group, leading to the release of the intron lariat and the ligation of the two exons. For details see main text.

b) The splicing procedure is regulated and accomplished by two main macromolecular protein complexes, the major spliceosome (U2- dependent), which is composed of mainly five small nuclear ribonucleoproteins U1, U2 U3, U4, U5 and U6, and the minor spliceosome (U12-dependent), composed of U11,U12, $U4_{atac}$, U5 and $U6_{atac}$. Illustration shows the identification of the main splicing sites SD, BP, poly-pyrimidine tract and SA, which are the binding sequences of U1, U2 and U2 auxiliary factors 65 kDa and 35 kDa, respectively. Regulation of splicing is modulated by cis-acting regulatory exonic and intronic enhancer (ESE and ISE) and silencer (ESS and ISS) elements, which constitute the binding sequences of the *trans*-acting factors. For details see main text. Pictograms at the bottom of the figure indicate the consensus sequences of the main splicing sites for the major (U2-type) and minor (U12-type) spliceosome in IUPAC code, with height proportions representing the relative frequency of each residue at the specific position. Y corresponds to cytosines C or U and R to purines A or G. Figure reproduced from [81]

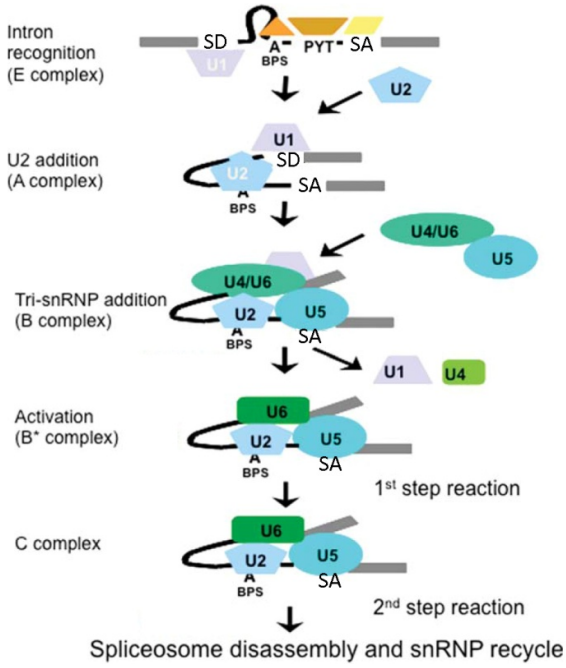


Figure 8: Spliceosome assembly activation and disassembly pathway.
Initially, introns are recognised by spliceosome where U1 snRNP, SF1 (orange triangle), U2AF 65kDa (brown rectangle) and U2AF 35kDa (yellow rectangle) bind to the 5' donor splice site (SD), branch point site (BPS), poly-pyrimidine tract (PYT) and 3' acceptor splice site (SA), respectively forming the E complex. Then, U2 snRNP replace the SF1, leading to the formation of A complex. Subsequently, the U4/U6.U5 tri-snRNP joins the complex and forms the B complex. Conformation changes and rearrangements lead to the release of U1 and U4 snRNP, which constitute the catalytically active B complex (B* complex), ready to complete the 1st step of splicing with the assistance of the DEAH-box RNA helicase Prp2 (Not shown). Then, the C complex is formed and further rearrangements lead to the 2nd catalytic step of splicing. Finally the spliceosome dissociates and the snRNPs are released to take part in additional rounds of splicing. Modified figure from [75]

In humans, exons are often small (~80% of the exons are < 200bp), surrounded and masked by large introns, which add an extra level of difficulty for distinguishing the conserved splice sequences, SD (AG/GURAGU), BPS (YNYURAY) and SA (NAG/G) (Figure 7) [81]. Beside of the context of the main splicing sites and regulation by the above-mentioned the flanking pre-mRNA regulatory element (ESE, ESS and ISE and ISS) and their *trans*-acting factors, proteome analysis of the human spliceosome indicates even greater complexity of regulation. A total of approximately 170 splicing-associated factors are involved at some stage during the splicing procedure, with individual intermediate complexes (e.g. A, B and C) containing up to ~125 proteins [76]. This illustrates the potentially cumulative effect of multiple factors that regulate splice site selection. More than 92–94% of the human protein-coding genes express multiple mRNA isoforms, as results of alternative splicing, contributing to the large diversity of proteins in higher eukaryotes, with wide-ranging implication in health and disease [73]. The most common mutations that change the normal splicing pattern are those that affect the *cis*-acting splicing core elements (SD, SA and BPS) or the flanking regulatory elements that regulate spliceosome recruitment (ESE, ESS, ISE and ISS) [81]. Examples of

these mutations are point mutations on *HBB* that lead to the formation of 3' aberrant SA sites (e.g. *HBB*$^{IVSI-110}$, in the following "IVSI-110"; HGVS name: HBB:c.93-21G>A), which change the splicing pattern and open reading frame (ORF), make the aberrant mRNA a target for NMD and in turn cause β-thalassaemia [82,83]. As additional disease examples, for Duchenne muscular dystrophy a splice-site mutation in dystrophin leads to loss of function by frameshifting in exon 51, for cystic fibrosis mutations of UG and of the U tract near the SA of cystic fibrosis transmembrane conductance regulator (*CFTR)* exon 9 cause disease of varying severity and for frontotemporal dementia with parkinsonism linked to chromosome 17, mutations of the 5'ss, ESE and ESS in microtubule-associated protein tau (*MAPT)* exon 10 are causative of disease [81]. In addition to *cis*-acting mutations, aberrant splicing can cause disease by mutations that affect *trans*-acting factors, such as components and associated factors of the spliceosome tri snRNP U4/6.U5 particle in the case of retinitis pigmentosa (pre-mRNA processing factor 3 (PRPF3), PRPF 4, PRPF 6, PRPF 8 and Bad Response to Refrigeration (BRR2)) or in colorectal carcinoma in which overexpression of the PRPF6 U5 – snRNP component promotes cancer cell proliferation [81].

1.8.2 *Defects in RNA stability and thalassaemia*

Expression and persistence of intact mRNA is important for sustained *HBB* expression. With a half-life of 10–24 h, the stability of normal globin mRNAs is in the middle of the range of the known half-lives of mRNAs, which vary between minutes and several days [60]. This stability allows the translation of globin chains to continue for up to three days after enucleation, permitting the high level of Hb production required for erythrocyte function [84]. This mRNA stability is regulated by the aforementioned *cis*-acting elements of the 5' and 3' UTR, the 5' cap structure, 3' poly-adenylation and regulatory sequences in the coding regions. A common feature in almost all eukaryotic mRNAs is mRNA circularization during translation, by linkage of the 5' cap structure and the 3' poly (A) tail through the poly (A) binding protein (PABP), which besides increased stability and protection of the mRNA termini against exonucleolytic degradation, also contributes to the initiation of translation and enhanced protein synthesis [60,73]. In addition, increased stability of HBB mRNA is also conferred by an HBB messenger ribonucleoprotein (mRNP) complex bound to the 14 nucleotide pyrimidine-rich track in the HBB mRNA 3' UTR, 34 nt downstream of the native termination codon. This mRNP complex contains one or more homologues of HBA poly(C) binding protein (αCP), which binds the pyrimidine-rich element (PRE) in the 3' UTR of HBA mRNA and contributes to the increased stability of the construct by inhibition of poly (A)-tail shortening and protection from erythroid endonuclease attack and 3'-to-5' exonucleolytic degradation [60]. In addition, a nucleolin RNA-binding factor binding site was identified in the 3' UTR of the HBB mRNA half-stem, opposite the PRE. This finding led to the

hypothesis of the HBB mRNA stabilization model, in which a stable stem loop is formed containing both elements, nucleolin binding site and β-PRE, with binding of nucleolin changing the confirmation of the loop and allowing the αCP protein to bind the PRE and stabilize the HBB mRNA [85].

As detailed in section 1.7.3, β-thalassaemia can be caused by mutations affecting promoter function or protein stability, direct or indirect interference with the functionality or stability of the HBB mRNA. One of the commonest results of such β-thalassaemia mutations is the formation of premature termination codons (PTCs), either by frameshift mutations (insertions/deletions), nonsense mutations or mutations affecting the correct splicing of HBB mRNA [56]. As protection against the potential toxicity of resultant truncated proteins, the nonsense-mediated decay (NMD) surveillance mechanism degrades abnormal mRNAs with PTCs. NMD recognises human HBB transcripts with a PTC downstream of codon 23 and approximately 50–55 nt upstream of the last exon-exon junction (Figure 10) [60]. In particular, PTC is identified by the presence of an exon junction complexes (EJC), 20–24 nucleotides upstream of the exon-exon junction sites, which are formed during the splicing procedure and remain bound to the mRNA until the 1st round of translation, when they are displaced by the translation machinery [86]. Briefly, the ribosome is able to displace the EJC during the elongation phase, however when it meets a PTC followed by an EJC at least at 50–54 nt away, then up-frameshift factor 1 (UPF1) and its associated SMG1 kinase bind to the translational release factors eRF1 and eRF3, forming the surveillance complex (SURF). In the meantime, UPF2 and UPF3b are bound to the downstream EJC, and its interaction with the SURF complex forms the decay-inducing (DECID) complex which leads to the phosphorylation of the UPF1 factor and release of the eFR1 and eFR3 factors. Then, SMG5, SMG6 and SMG7 factors are recruited to the complex with other mRNA degradation factors, which eventually leads to mRNA decay, de-capping and 5’ to 3 exoribonucleases degradation by XRN1, deadenylation and 3’-to-5’ exonucleolytic degradation by exosomes (Figure 9) [60,86]

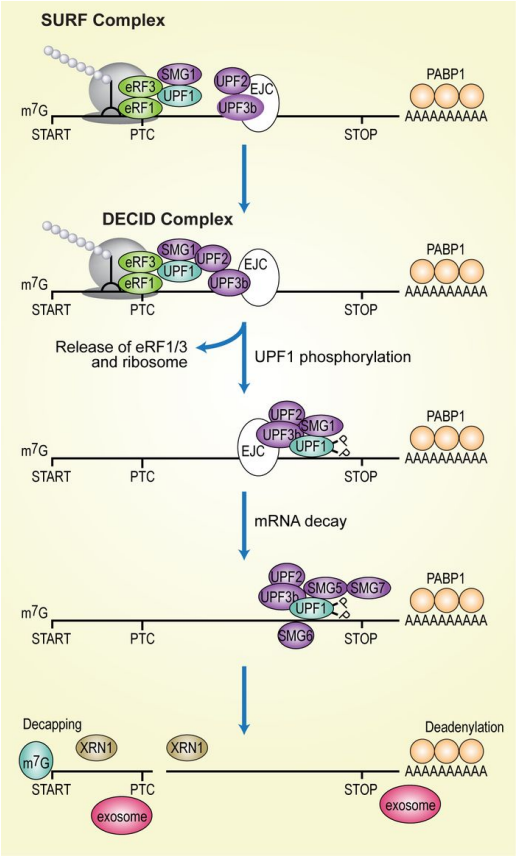


Figure 9: Simplified non-sense-mediated decay (NMD) pathway.

In the presence of a premature termination codon (PTC), the up-frame 1 factor and its associate kinase SMG1 binds the eukaryotic release factors (eRF1 and eRF3) and form the surveillance complex (SURF). Interaction of the SURF complex with the UPF2 and 3 factors bound on the adjacent exon junction complex (EJC) triggers the formation of the decay-inducing complex (DECID), leading to the phosphorylation of UPF1 by SGM1, release of the eRF1/3 factors and ribosome from the complex and the recruitment of SMG5, SMG6 and SMG7 and other mRNA degradation factors, which eventually cause mRNA degradation. m7G: 5' cap, PABP1 polyadenylation binding protein 1, XRN1 5'-to-3' exoribonuclease 1. Figure reproduced from [86].

In the case of β-thalassaemia and in agreement with the 50–54 nucleotide rule, they have shown that β-thalassaemia mutations causing PTC in the 3' end of the exon 1 and within the 5' two-thirds of exon 2 (e.g. 26, 36, 39, 60/61, 75 and 89) are NMD-competent, whereas those in the 3' end of exon 2 and exon 3 are resistant to NMD. In fact, mutations that cause PTC late in the 3' end of exon 3 are able to produce truncated HBB chain variants which, as described in section 1.7.3, precipitate and are therefore characterised as dominant mutations (e.g. PTC 121, 127 and 141). While PTCs in the 3' end of exon 2 and 5' end of exon 3 are resistant to NMD, the short truncated HBB chains are effectively degraded via proteolysis (e.g. PTC 88, 101 and 114). PTCs at the 5' end of exon 1 are similarly NMD resistant and produce short peptides, which are efficiently degraded (e.g. PTC 5, 15 or 17) [60]. Summarising, transcripts with PTCs outside the limits designated in Figure 10, escape the NMD surveillance pathway, and are translated to truncated proteins, which are subsequently degraded via proteolysis [56,60,87]. While β-thalassaemia transcripts under NMD control are generally

recessively inherited, those with PTCs beyond the upper NMD recognition border may produce potentially toxic non-functional HBB proteins, long enough to escape proteolysis, and may cause β-thalassaemia intermedia in heterozygotes with dominant inheritance [56,60,61,88]. While NMD is therefore generally beneficial, it might also remove truncated and partially functional products that would lead to a milder phenotype in the absence of NMD degradation [89]. Overall, the NMD pathway plays an essential protective role, contributes to observed phenotypes in β-thalassaemia heterozygotes and is a co-determinant of disease severity in β-thalassaemia patients with different mutations.

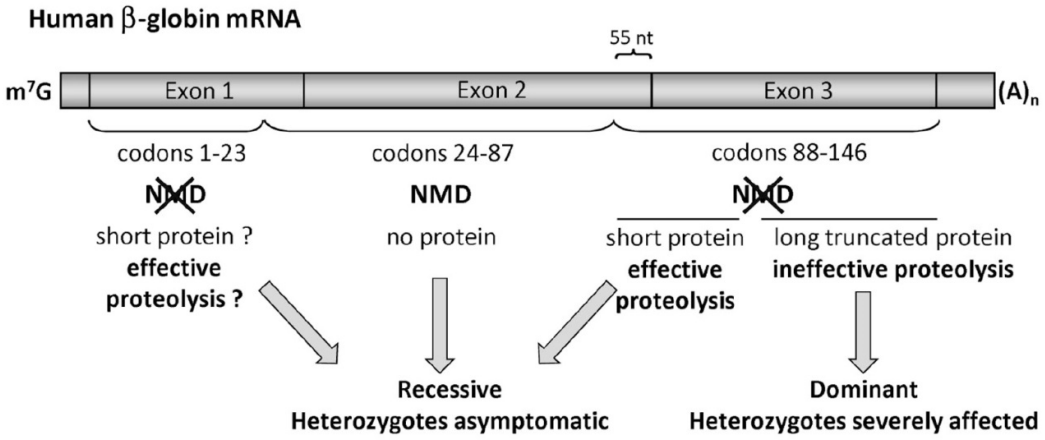


Figure 10: Modulation of β-thalassaemia phenotype by nonsense-mediated (mRNA) decay
The map indicates the regions of the HBB mRNA which are NMD-resistant and NMD-sensitive. PTCs near start codon (1-23 codons) are resistant to NMD, and heterozygotes are asymptomatic since translated short HBB globin peptides along with the excess of HBA chains are effectively degraded. Transcripts with PTCs downstream of the 23 codon and at least 55 nucleotides (nt) upstream from the last exon-exon junction site, are sensitive to NMD and heterozygotes are also asymptomatic. If the PTCs are located closer to the last exon-exon junction site (55 nt <) or in the 5' end of exon 3, then transcripts escape the NMD leading to the production of small truncated HBB chains which are effectively degraded along with the excess of HBA chains, therefore heterozygotes are asymptomatic. Transcripts with PTCs further downstream in exon 3 are NMD-resistant, leading to the production of toxic truncated non-functional HBB chains long enough to escape proteolysis, and cause β-thalassaemia intermedia in heterozygotes with dominant inheritance. m7G, 5'-cap; NMD, nonsense-mediated decay; (A) n, poly (A) tail. Figure reproduced from [60].

1.8.3 *IVSI-110 mutation and other splice-site mutations*

One of the commonest causes for thalassaemia is alterations of HBB pre-mRNA splicing [56,90]. Mutations affecting the invariant SA and SD sites, GT and AG, respectively, completely abolish correct splicing and are designated as β^0 mutations, which lead to thalassaemia major, whereas those affecting adjacent conserved sequences allow a proportion of pre-mRNA to be correctly spliced, leading to the synthesis of functional HBB chains and therefore designated as β^+ mutations. Homozygotes for β^+ mutation usually cause thalassaemia intermedia, which, as explained in section 1.7.3.2, covers a vast range of phenotypes, from a pathophysiology bordering on constitutively transfusion-dependent thalassaemia major to asymptomatic conditions, which might be mistaken for β-thalassaemia trait [56,61].

The IVSI-110 mutation (HGVS: HBB:c.93-21G>A) follows a recessive inheritance pattern and is one of the commonest β-thalassaemia mutations worldwide (Figure 11), affecting ~76% of β-thalassaemia carriers in Cyprus, according to the online database IthaMaps (http://www.ithanet.eu/db/ithamaps) [54]. The (G>A) point mutation creates an abnormal 3' SA site, 19 nucleotides upstream of the original intron-1 SA site, leading to the incorporation of the 19-nt sequence and an in-frame PTC into the mRNA, which is thus efficiently targeted by NMD [82,83]. Although ~10–20% of transcripts are correctly spliced and give up to 10% of normal Hb, so that IVSI-110 is categorized as β^+, the mutation causes a severe form of thalassaemia major, in which patients require regular blood transfusion and chelation therapy to survive [56,90]. Analysis of the strength of each SA sites (aberrant and normal), using the Shapiro and Senapathy matrix, indicate a higher score for the normal SA site (84.5) than the aberrant site (80.1), which contradicts the observed preference of the spliceosome for the aberrant SA site [90]. Analysis of both sites by the exonic splice enhancer finder algorithm (ESEfinder) indicates that the IVSI-110 (G>A) mutation disrupts the binding motif of the SRp40 protein factor, which may disrupt activity of SR splice enhancer for the normal SA site [90]. We can also speculate that disruption of this motif could reduce the efficiency of spliceosome assembly and recruitment to the normal SA site, therefore favouring the mutant SA site. In both cases the spliceosome use the 5' SD site and normal branch site [91,92]. In addition, a pertaining mouse model was developed by Vadolas *et al.*, harbouring the IVSI-110-mutant *HBB* locus in an $Hbb^{th3/+}$ background, displaying the same reduction of HBB protein synthesis observed in IVSI-110 patients [90]. Other, less common splice-site mutations, such as IVSII-745 (HGVS name: HBB:c.316-106C>G); with highest carrier frequency in Jordan 14.2% and as the second-most common *HBB* mutation in Cyprus with 3.79%, see IthaMaps (http://www.ithanet.eu/db/ithamaps), cause thalassaemia in a similar fashion, are subjected to NMD in line with the rules set out above and may benefit from research and models developed for IVSI-110 [93].

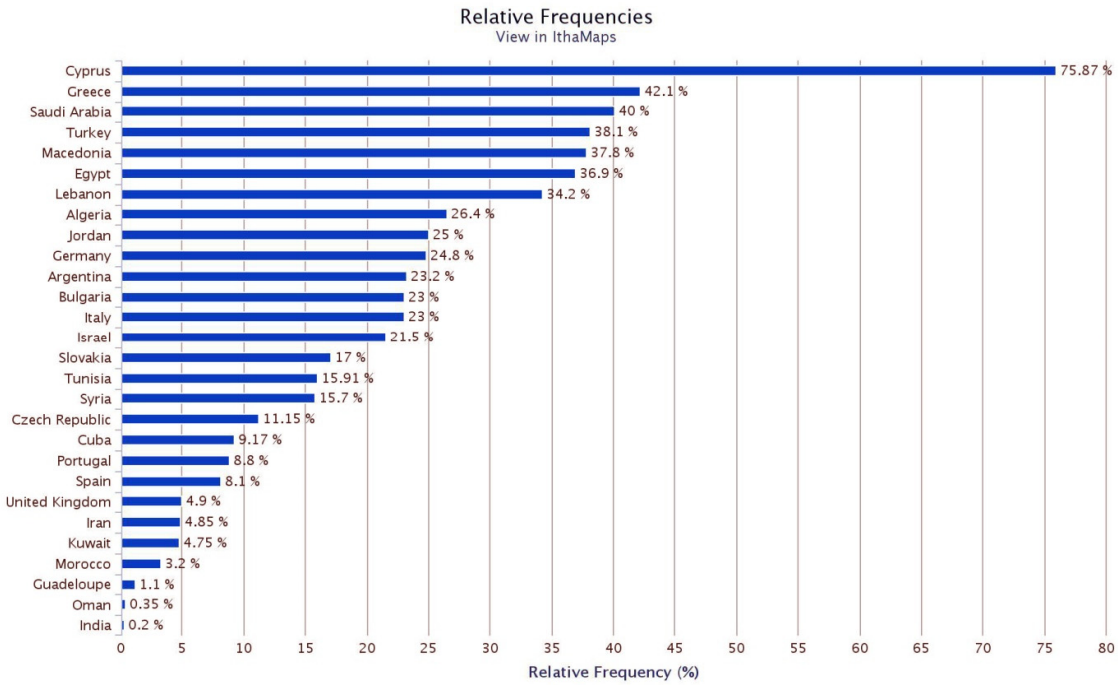


Figure 11: **Relative carrier frequencies for the IVSI-110 mutation world-wide.**
Data obtained from the IthaMaps web data base http://www.ithanet.eu/db/ithamaps described in [54]

The first IVSI-110-specific treatment used IVSI-110 HeLa cells and antisense oligonucleotides (ASON) matching the normal BPS or the aberrant 3' SA site, in order to correct HBB pre-mRNA splicing. While targeting the BPS forced the use of the normal SA site and restored normal splicing more efficiently in that study, an independent effort targeting the 3' SA site by antisense RNAs (asRNA) recently gave 50% efficiency in restoring normal splicing [91,94]. When a modified U7 snRNA carrying an asRNA for an IVSII-745 splice enhancer sequence was applied to IVSII-745/IVSII-1 erythroid progenitor cells using an integrating lentiviral vector (LV), correct splicing was increased 25-fold compared to controls [95]. The latter approach points the way away from transient approaches that would require chronic treatment, towards a repair of defective splicing, here at the RNA level, by stable and long-term therapy for thalassaemia [96].

1.9 Therapeutic approaches to β-thalassaemia

1.9.1 *Management and current treatments*

Today the only available curative treatment for thalassaemia is allogeneic haematopoietic stem cell transplantation (allo-HSCT) from BM or mobilised $CD34^+$ cells. Allo-HSCT, however, is limited to the fraction of patients with histocompatible donors, is poorly tolerated in adults, and holds a 5–10% mortality rate associated with the procedure and additional mortality or morbidity through subsequent immune complications, such as graft-versus-host disease (GvHD) [66,97]. In general,

approximately only 25–30% of thalassaemia patients have a matched sibling donor [57]. In addition to BM or mobilised $CD34^+$ cells, $CD34^+$ from the cord-blood of matched related donors constitute an additional attractive stem cell graft source for thalassaemic patients, with high probability of a successful cure and low risk of developing GvHD [98–100]. For leading groups and based on refined methodology for the conditioning of recipients and for the analysis of compatibility, even matched unrelated or haploidentical donors of HSCs give good efficiencies and disease-free survival after transplantation [101]. In all instances, GvHD is one of the main potential complications post-transplantation, so that even after successful transplantation approximately 5–8% of the treated patients may develop chronic GvHD disease [57]. The main treatment of the disease is immunosuppressive medicines, which are also associated with major risk of complications or death by bacterial, fungal and viral infections.

The most extensively used approach for managing β-thalassaemia is life-long regular blood transfusion, in order to correct the anaemia, but without ameliorating the toxic complications secondary to iron-overload. Patients with the most severe forms of thalassaemia major die within the first 2 years if not treated with regular blood transfusions [61]. Requirement and frequencies of blood transfusions varies based on severity of thalassaemia. The main goal of blood transfusion is to increase the levels of Hb from the <8 g/dL seen in severe forms of anaemia to 13–14 g/dL post transfusion and to maintain the levels above 90–100 g/L at all times [29]. However, as a result of regular blood transfusions and the increased gastrointestinal iron absorption, patients start to present the complications of iron overload, such as cardiomyopathy, liver fibrosis and endocrine dysfunction. Therefore, continual and cautious monitoring of iron concentration should be performed with serial measurements of serum ferritin, which has been found to correlate with body iron stores [57]. In order to ameliorate the complication of iron overload, transfusional iron is removed using chelating drugs, such as deferoxamine, deferiprone and most recently deferasirox, aiming to maintain the concentrations of iron in the body to non-toxic levels [102]. Chelation monotherapy or chelators in combination are able to effectively reduce excess iron, allowing regular and prolonged blood transfusions and reducing the morbidity and mortality of β-thalassaemia. Of note, iron chelation therapy might also be necessary for patients that are transfusion independent, such as in the cases of mild thalassaemia intermedia, since iron overload can be developed due to ineffective erythropoiesis, which triggers increased iron absorption by inhibiting the synthesis of hepcidin [66]. Combinations of regular blood transfusions and iron chelation have extended the life expectancy of thalassaemic patients beyond the age of 40 [57].

According to a World Health Organisation (WHO) estimate for 2006, more than 330 000 affected infants are born per year, from which approximately 83% suffer from SCD and 17% from thalassaemia [55]. In addition, globally ~7% are carriers for haemoglobinopathies which ranks the group of disorders amongst the most common monogenic disorders, making it one of the major world health problems. Lifetime treatment of patients with β- thalassaemia major in the United Kingdom has been estimated to between £188 000 – £226 000 [102]. Treatment of such patients could have a major impact on national health budgets of a developing economy, as has occurred in Cyprus and Sardinia in 1960 [29]. Therefore, the WHO in an attempt to address and constrain the growing problem of haemoglobinopathies, proposed prevention and management programmes. This includes public education, population screening and counselling programs, and the development of prenatal diagnosis and pre-implantation genetic diagnosis, which in combination with chronic blood transfusions and chelations constitute the global standard for disease control and management [48,55,102]. Differences in culture and religion among countries create problems for the implementation of prevention and management programmes, therefore customisation for the specific societal context, public education and long-term discussion between communities and specialised doctors are essential for the successful establishment of these programmes [48].

1.9.2 *Novel therapies based on small molecules*

Inspired by the therapeutic effect of *HBG* expression in HPFH adults with β-thalassaemia (see above), chemical induction of *HBG* expression has been pursued as a therapeutic option for decades. Over the years, countless HbF-inducing agents have been identified, for many of which the precise molecular mechanism or targets are still unclear [103]. HbF inducers, such as DNA demethylating agents (5-azacytine, 5-aza-2-deoxycytidine, also known as decitabine), cytostatic agents (hydroxyurea, HU) and histone deacetylase inhibitors (HDAC) (butyrate) and short-chain fatty acids, have found clinical application for SCD, but for β-thalassaemia are still plagued by inconsistent efficiency in different patients and by a generally high level of toxicity. One of the major concerns over using these agents is the long-term side effects, including carcinogenesis potential (e.g. 5-azacytidine). Compounds which are able to alter the chromatin stability in a non-specific manner, such as demethylating agents or deacetylase inhibitors, may also activate or silence other genes, increasing the risk for carcinogenesis. In particular, concern over increased carcinogenicity of 5-azacytidine has prevented its use for human treatment and has led to the development of safer and more efficient compounds at lower concentrations, such as decitabine [104,105]. Another drawback of chemical HbF-inducing agents is the requirement of chronic application over the lifetime of the patient. Depending on the reagent, reapplication might need to be frequent in order to sustain

sufficient HbF levels, owing e.g. to low efficiency or short half-life of the compound *in vivo* (arginine butyrate). The nature of the compounds might also limit the method of administration to either oral administration (e.g. HU and butyrate alternatives: Sodium phenylbutyrate and isoputyramide) or intravenous transfusions (arginine butyrate, 5-azacytine and decitabine) [106].

The alternative approach of targeting PTC mutations by PTC-read-through agents is as yet in its infancy but promising for nonsense mutations, such as the frequent codon 39 β^0 mutation (HGVS name: HBB: c.118C>T) for β-thalassaemia. A recent study showed that approximately 11% of all described gene lesions responsible for inherited human diseases are due to non-sense mutations [107]. The aim of this approach is the decrease of the ribosome accuracy to recognize PTC and proceeding to the translation of a full length and fully or partially functional protein. This PTC-read-through agents are split in two categories, the aminoglycoside antibiotics, such as geneticin (G418) and gentamicin and the non-aminoglycoside compounds (such as PTC124 (Ataluren), RTC13 and RTC14) [108]. Compounds of the former group have been shown to efficiently read through PTC and lead to the synthesis of a full-length functional protein, such as in the cases of *in vitro* and *in vivo* mouse models for DMD, cystic fibrosis, haemophilia and β-thalassaemia. However, these drugs also presented severe nephrotoxic or ototoxic side-effects, which would make long-term therapeutic application of these drug untenable [108–111]. In particular, in the case of thalassaemia, aminoglycosides showed increased HBB chain synthesis and accumulation of HbA in erythroid precursor cells from cd39 (HGVS name: HBB:c.118C>T) β^0-thalassaemic patients with the use of G418 [112]. The non-aminoglycoside compounds on the other hand lack antibiotic properties of the aminoglycosides, rendering them safer and thus more suitable for pre-clinical trials. RTC13 and RTC14 were shown to induce read-through of PTCs in ataxia-telangiectasia mutated (*ATM*) and *DMD* genes [113]. Another example of non-aminoglycoside compounds is amlexanox, which in addition to the PTC-read-through property also acts as NMD inhibitor and was shown to increase the levels of mRNA containing nonsense mutations by two- to five-fold without affecting the natural NMD targets or observed cytotoxicity in mammalian cells [111]. Global interference with NMD pathways, however, has a high risks of side effects, since besides its surveillance and protective role against the translation of aberrant mRNA, NMD is involved in the regulation of expression of ~10% of the mammalian transcriptome and is essential e.g. during embryonic development [111]. Another approach specific to stop-codon mutations reaches into the realm of GT and is the use of nonsense suppressor tRNAs, which are modified tRNA derivatives carrying a specific amino acid and a modified anticodon specific to the PTC. This approach leads to the incorporation of the specific amino acid at the site of the PTC leading to the production of full-length protein with full or partial activity. Besides the common requirement for efficient delivery that is shared by all GT approaches, nonsense-tRNA suppressors

also face the difficulty of potentially low PTC suppression efficiency, first due to low aminoacylation in competition with normal tRNA and second due to dependence of PTC suppression on PTC position and sequence context. Most importantly, the approach would be marred by side effects owing to suppression of all normal occurrences of the same termination codon and the production of deleterious proteins [111].

In summary, small-molecule-based drugs exhibit promising results but after decades of development still need further optimisation of specificity and tolerability for safe long-term application in humans. The need for chronic lifelong application represents an acceptable drawback in practice, while toxicity and corresponding long-term side-effects because of unspecific action are an inherent problem of all approaches presented in this section and fundamental impediments to their clinical translation. It has become increasingly clear for any type of established or novel treatment for β-thalassaemia that treatment needs to be customised if not personalised in order to achieve good efficiency and tolerability, possibly by inclusion of pharmacogenomics analyses and certainly by taking specific primary mutations and modifying factors into account [49,114,115].

1.9.3 *Gene therapy for the correction of β-thalassaemia using LV-based approaches*

Where conservative disease management by transfusion and chelation is ineffective, alternative choices are all but inviting. While palliative small-drug approaches to β-thalassaemia treatment display low efficiency and high toxicity, conventional stem-cell-based approaches carry the risk of immunological complications. However, most limitations and risks associated with allo-HSCT would be removed by genetic correction and re-implantation of autologous cells instead. In this vein, β-thalassaemia as a monogenic disorder of the haematopoietic system is an ideal target for GT by gene addition with LVs. These have the ability to infect non-dividing cells and hence quiescent HSCs, can export and package long (~8-kb), unspliced RNAs for therapeutic use and through stable integration into the host genome may be curative, with a lower risk of transgene rearrangements and of insertional mutagenesis than γ-retroviral vectors [116]. In particular, LVs have a safer integration profile than γ-retrovirus, with the former to favour the integration throughout the transcription unit in gene dense regions, whereas the latter to shows a preference for integration at the transcription start site of cellular genes, promoters and 5' regulator regions [117,118].

The main principle of additive GT for thalassaemia is the delivery and expression of functional globin transgenes at therapeutic levels, which for clinical application needs to show position independence and permanence of high-level erythroid-specific expression at minimum VCN, in order to minimise the chance of insertional mutagenesis, genome instability and toxicity of viral- or vector-encoded

proteins [66,118–120]. Safety of the use of LV in to GT was increased with a self-inactivating design (SIN), in which promoter/enhancer sequences in the U3 region of the 3' long terminal region (LTR) was deleted. This deletion with the reverse transcription of the RNA genome to complementary DNA (cDNA) is transferred to the 5'LTR, which self-inactivates the provirus and leave active only the promoter and regulatory element of the transgene. In addition, this design minimise the transactivation of neighbouring cellular promoters due the LV strong LTR enhancers as also and the long-term silencing of the therapeutic gene due to methylation of the viral promoter [118]. Importantly, it reduced the risk of generation of replication-competent recombinant LV in the producer and target cells. The large capacity of LVs which allows the transfer of the *HBB* therapeutic transgene and the regulatory elements (βLCR) essential for high and erythroid-specific expression and the ability to transduced non-dividing cells (i.e. LT-HSC), prompted their use in the GT of haemoglobinopathies and in particular thalassaemia and SCD [118]. Thalassaemic murine models ($Hbb^{th1–4}$) with partial or complete deletions of the mouse *HBB* genes, β^{major} and β^{minor}, have already been treated with LVs harbouring suitably regulated *HBB*-or *HBG*- transgenes [114,121–128].

1.9.3.1 *Pre-clinical studies*

A large number of independent studies, too numerous to detail here and recently reviewed elsewhere, have over the last decade pushed the field of LV-based GT of β-thalassaemia to the execution of the first clinical trials [39,66]. Exemplary work was performed by May *et al*., who showed that incorporating long HS2–HS4 LCR elements instead of merely the corresponding cores vastly improved transgene expression, allowing the correction of anaemia in $Hbb^{th3/+}$ mice with a VCN below 1.0 (in nucleated peripheral blood cells) and in the $Hbb^{th3/+}$-based surgical thalassaemia major model to the level of thalassaemia intermedia, so that a SIN-modified version of the original TNS9 proceeded to phase-I human clinical trials, albeit without achieving transfusion independence for its trial subjects (see below) [39,126,127,129]. A similar vector design employed by Pawliuk *et al.* was therapeutic only at an average VCN of 3 in either a murine SCD (in nucleated peripheral blood cells) or mild thalassaemia intermedia (in bone marrow, spleen and thymus cells) background [121,130]. Inclusion of an amino-acid substitution in the *HBB* transgene (T87Q) added anti-sickling properties to the vector-encoded product in the context of SCD and would eventually also allow its discrete detection and quantification during the follow-up phase of the first clinical GT trial for β-thalassaemia (see below) [130,131].

Attempts at improving long-term and position-independent expression were pioneered by the group of Malik in vectors akin to those employed above by the groups of Sadelain and Leboulch. To this end she used a variety of versions of the 1.2-kb chicken *HBB* hypersensitive site 4 (cHS4) insulator for

inclusion in the provirus, with the 400-bp core region not retaining the necessary activity, with the full-length version reducing viral titres to a levels unsuitable for clinical trials and with a partially active twin core adopted by the Leboulch group in the LentiGlobin HPV569 vector for the first clinical β-thalassaemia GT trial (See below).

In a most recent development and based on the TNS9 design, the Rivella group have developed the AnkT9W vector, which carries the erythroid-specific ankyrin 5' HS barrier insulator, corrected murine thalassaemia at significantly lower VCN (in erythroid progenitor and $CD34^+$ cells) than the base vector and reached carrier levels of HBB for $\beta^{0/+}$, $\beta^{+/+}$ and, for higher VCNs, $\beta^{0/0}$ samples from patients [114].

The GLOBE vector, on which my own work is based, has been developed by Michael Antoniou at KCL, UK, and exceptionally carries an LCR reduced to elements HS2 and HS3 to drive expression of a modified human *HBB* gene under control of the *HBB* promoter. In particular, *HBB* transgene with the mini βLCR was inserted in SIN LV (400 bp deletion of HIV U3 LTR region) in reverse orientation compared to the LV reverse transcription orientation. In addition, the *HBB* transgene size was further reduced with a deletion in Intron 2, reducing its size to 257 bp and retaining the 3' *HBB* enhancer [132]. The whole design of the GLOBE provide the efficient expression of the *HBB* transgene and the production of high titres of stable LVs, requirements necessary for large scale experiments and its utilisation into the clinic [132]. In studies conducted by the Ferrari group, GLOBE was able to correct the phenotype of murine thalassaemia intermedia ($Hbb^{th3/+}$) at a VCN of 1 (in BM HSCs) and with transduction of only 30–50% of HSCs, while correcting a surgical thalassaemia major model either to normal or thalassaemia-intermedia phenotype at a VCN >3 (in BM HSCs) [132]. The study alongside others indicated in GT-treated mice what is also known from allo-HSCT: that owing to a survival advantage of non-thalassaemic erythrocytes, a relatively low (20–30%) BM chimerism of corrected or healthy donor HSCs may be sufficient for the resolution of thalassaemia [128,133]. These findings suggest that partial myeloablation of patients before re-infusion of genetically corrected HSCs may be sufficient for GT-based correction, thus avoiding the dangers, such as increased toxicity and the risks of GvHD and organ failure, associated with full myeloablation [66,134]. Subsequent to the correction of murine thalassaemia mice, Roselli *et al*., demonstrate the therapeutic potential of GLOBE on thalassaemia major patient-derived $CD34^+$ cells from BM, carrying different mutations. In particular, they showed that GLOBE was able to correct the thalassaemic phenotype by increasing the production of HbA at carrier's levels at low VCN (VCN ~1.6 in $CD34^+$ cells) without showing integration preference to proto-oncogene, tumour-suppressors or cell-cycle related genes [115]. The therapeutic benefit of *HBB* transgene expression rescue cells from apoptosis and ineffective

erythropoiesis, in which more mature erythroblasts were detected in transduced samples compared to the untransduced [115]. In an attempt to improve GLOBE LV, Miccio *et al.*, subsequently inserted the HS2 enhancer of the *GATA1* gene into the 3' SIN lentiviral LTR of GLOBE. In the provirus, the enhancer was correspondingly present in both LTRs and protected the therapeutic *HBB* transgene from chromatin silencing in order to maintain higher levels of expression, so that correction of $Hbb^{th3/+}$ thalassaemic mice was achieved at lower VCN than with the original vector [135].

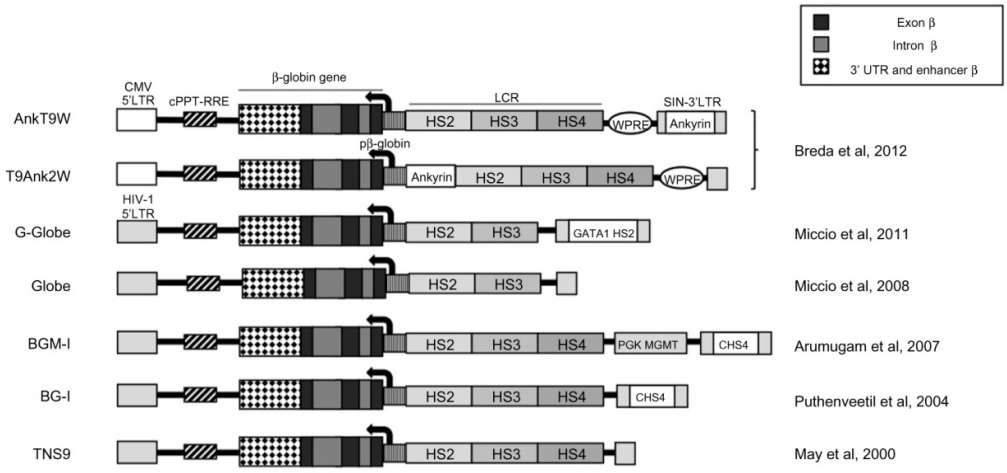


Figure 12: Lentiviral vectors expressing the HBB transgene.
Ankyrin, ankyrin insulator; cHS4, chicken *HBB* hypersensitive-site-4 insulator; CMV, Cytomegalovirus; GATA1 HS2, globin transcription factor 1 hypersensitive-site-2 enhancer; HIV-1, human immunodeficiency virus; HS, hypersensitive site; LCR, locus control region; LTR, long terminal repeat; pβ-globin, *HBB* promoter; PGK MGMT, murine phosphoglycerate kinase-1 promoter and O-6-methylguanine DNA methyltransferase gene; SIN, self-inactivating; UTR, untranslated region; WPRE, woodchuck posttranscriptional regulatory element; cPPT, central polypurine track; RRE, Rev response element. Reproduced from [39]

1.9.3.2 *Clinical trial(s)*

Up to now twenty-eight patients were treated with *HBB* lentiviral vectors in three clinical trials for β-thalassaemia [131]. The first human thalassaemia GT clinical trial was performed by the Leboulch group with the HBB^{T87Q} modified therapeutic vector (LentiGlobin HPV569) [136–138]. LentiGlobin HPV569 holds the HS2-4 regulatory elements of the βLCR and the anti-sickling HBB^{T87Q} transgene where two core elements of cHS4 insulators (2x 250bp) were cloned at the site of the U3 deleted region of the 3' LTR. These elements are considered to provide a position independent expression of the HBB^{T87Q} transgene [136]. Initially three patients with transfusion-dependent β^0/β^E thalassaemia were treated with the HPV569 vector. Of these, one failed to accept the transplant after complete myeloablation and received autologous back-up cells and the third even though engraftment of the transduced cells was successful, remained transfusion-dependent, likely owing to the low VCN in the transduced cell population (VCN ~0.3) and in nucleated cells in peripheral blood post-engraftment (VCN 0.0016 in neutrophils), so that vector-derived Hb constituted only 5% of total Hb 2 years post-engraftment [39,139]. The 2nd patient of this initial trial reached transfusion-independence 12 months after treatment

and continues to date (8 years after), showing an apparently stabilised clonal expansion of one cellular clone with a rearranged proviral insert and a constant and equal contribution of HbE, HbF and vector-derived HbA^{T87Q} to total Hb (8 g/dL) from year 2 to 8 post-engraftment [131,137,139]. Therefore the high expression of the endogenous Hbs HbF and HbE appears essential for the patient to achieve and maintain transfusion-independence. Quantification of the VCN in peripheral neutrophils was 0.23 [139]. Integration site analysis of the predominant clone showed that the vector was integrated in intron 3 of the high mobility group AT-Hook 2 (HMGA2) gene and that one of the two cHS4 elements was lost, probably due to a recombination event, resulting in activation of the *HMGA2* promoter by the vector-encoded βLCR elements and in increased expression of the gene in the erythroid cells. Furthermore, a cryptic SA site located in the cHS4 core element of the 5' LTR lead to abnormal splicing of the HMGA2 mRNA, which also drew on RNA-cleavage and polyadenylation sites within the R region of the 5' LTR to produce a truncated HMGA2 protein with 3 exons. Exclusion of exon 5 from the mature mRNA removed also the Let7 miRNA binding site, which acts as a regulatory element of *HMGA2* expression, therefore increasing stability and excess of the truncated HMGA2 protein in erythroid cells (~10 000-fold) [39]. *HMGA2* is associated with tumour metastasis and proliferation [140], so that clonal dominance of this clone was a concern until it peaked 4 years post-treatment at 22% of nucleated cells and dropped to 6.8% a year later, with continued transfusion independence of the treated-patient [39,139] Stable contribution of vector-derived HBB^{T87Q} chains throughout the later post-transplantation period indicated that the therapeutic effect is not dependent of the specific clone [131]. Taken together, clonal expansion, low vector-derived gene expression, vector rearrangements and slow progress to transfusion independence despite the full myeloablation of the patient, demonstrated the need for improved protocols and vectors for successful and safe future clinical trials. HPV569 design has come into question independently, with experiments in mouse demonstrating that most of the integrated vectors lost one of the two cHS4 core elements, due to recombination/rearrangements events, diminishing its protective role against the activation of adjacent genes [141], which moreover is dependent on the cHS4 integration site in the genome [142]. Therefore, the Leboulch group proceeded to the optimization of the LentiGlobin HPV569 vector to produce the BB305 LV. Key improvements were the removal of the cHS4 elements and replacement of the 5' LTR Tat-dependent U3 promoter/enhancer with a CMV promoter and enhancer increasing stability and safety, and vector titre and transduction efficiency, respectively. The provirus of the optimised vector, LentiGlobin BB305, is identical to that of its predecessor HPV569, apart from the absence of the two cHS4 insulator elements. Comparison of both vectors showed 3–4-fold titre and 2–3-fold transduction efficiency for BB305 compared to HPV569 on human $CD34^+$ cells, while both vectors corrected the mild-thalassaemic phenotype of $Hbb^{th1/th1}$ mice

with equivalent safety, i.e. without association with insertional oncogenesis [143]. After these encouraging results they moved to a new round of clinical trials with the BB305 vector. These trials started in France (HGB-205) for β-thalassaemia and SCD, where up to now four β-thalassaemic (three major β^0/β^E and one homozygote for IVSI-110) and one SCD patient have been treated with encouraging results especially for the latter[131]. In particular, the SCD patient is expressing a total of ~51.5% of anti-sickling Hb (48% HbAT87Q, 1.8% HbF, 1.7% HbA2) 9 months post-transplantation, whereas the two β^0/β^E thalassaemic patients with VCN 1.5 and 2.1 (*ex-vivo* on pooled colony-forming progenitors) remain transfusion independent for at least 15 months post-treatment [39,144,145], whereas the most recently treated fourth patient, an IVSI-110 homozygous thalassaemic patient who is incorrectly described as a β^0/β^0 patient in corresponding publications, has also become transfusion independent [131,139,145,146]. Further international clinical trials (USA, Thailand and Australia (HGB-204) and USA (HGB-206)), with BB305 are now in progress, also with encouraging results [131].

The Sadelain group has also been approved as the first US clinical trial for β-thalassaemia using the TNS9.3.55, which holds the cHS4 insulator with minor unpublished modifications compared to the original TNS9 LV [125,126]. A pilot-trial was performed in order to assess the safety and efficiency of $hCD34^+$ mobilisation from β-thalassaemia patients using granulocyte colony-stimulating factor (G-CSF), and the efficiency of the clinical-grade TNS9.3.55 vector used under clinical conditions. The study achieved a high level of mobilisation of more than $8x10^6$ $CD34^+$ cells/kg with minimal side-effects. In addition, $hCD34^+$ cells of three different patients were efficiently transduced, giving an average VCN of 0.53. Normalisation of the expression of vector-derived HBB chains based on the VCN indicated that transgene expression per vector copy reached that of normal hemizygous protein levels (70-100%) [147]. This supported the hypothesis that a VCN of 1 would be sufficient to correct the thalassaemia intermedia phenotype and possibly thalassaemia major. In addition, transplantation of the transduced cells into NOD/SCID/IL2rγ^{null} mice indicated a stable VCN (in BM HPCs) and high expression of the *HBB* transgene 6 months post-transplantation [147]. Data from this study constituted the basis of protocols that were followed and applied in the subsequent clinical trial of the TNS9.33.5 LV [147]. In the clinical trial, four out of five originally registered patients have been treated up to now, without reaching transfusion independence 12 months post treatment and with only one showing a significant decrease of transfusion requirements. In contrast to LentiGlobin trials, where full myeloablation of the patients is performed (14 mg/kg busulfan), the first three β^0/β^+ patients for the TNS9.3.55 LV underwent mild conditioning (8 mg/kg busulfan), and the percentage of transduced mononuclear cells in peripheral blood increased from initially 1% to 7–9% [39]. Conditioning of the fourth β^0/β^0 patient was increased to 14 mg/kg busulfan [131] in an effort to improve phenotypic correction, with results pending.

Finally, preliminary results of clinical trials, performed by Ferrari's group, using the GLOBE LV as the third GT platform in β-thalassaemia clinical trials are highly encouraging. The phase-1/2 clinical trial uses three patient cohorts: a) adults (≥18 years) 3 subjects, b) elderly children (8-17 years) 3 subjects and c) younger children (3-7 years) 4 subjects (Clinical trials.gov identifier: NCT02453477) (https://clinicaltrials.gov/ct2/show/NCT02453477). Conditioning for the first two cohorts will be performed with treosulfan and thiotepa and with busulfan for the third. Collection of autologous HSC for adults and elderly children will be collected from mobilised peripheral blood (PB) using combination of granulocyte-colony stimulating factor (G-CSF) and plerixafor (PF) mobilizing agents and from BM in young children when mobilisation is not possible. At this point in time, the study has concluded recruitment for the first two cohorts and has preliminary results for the first. The 1st patient, 6 months post-GT was transfusion independent for 5 months with average VCN 0.7 (in BM HPCs), the 2nd patient three month post-GT have VCN 0.9 (in BM HPCs) whereas for the 3rd patient no data were presented as is still too early (59 days post-treatment). However, overall patients showed good tolerability to the procedure in which good yields of hCD34^{+} cells were obtained with the combinational G-CSF and PF mobilisation [148].

1.9.4 *RNA Interference*

RNA interference (RNAi) is a conserved mechanism for post-transcriptional gene regulation, in which short RNA molecules convey sequence-specificity for the translational inhibition (by miRNAs) or enzymatic degradation (by short interfering RNAs, siRNAs) of target RNAs [149].

1.9.4.1 *RNAi mechanism*

In mammals, miRNAs have an essential role in many developmental and cellular processes and regulate the expression of more than 30% of the human transcriptome with more than 1000 human miRNAs [150]. miRNAs are endogenously expressed as primary miRNA (pri-microRNA) by RNA pol2 and form distinctive imperfect hairpin structures. These structures are recognised and processed by the nuclear RNase-III-like endonuclease Drosha and its co-factor DGGR8 to ~70-nucleotide (nt) dsRNA molecules known as pre-miRNAs. Subsequently, the pre-miRNAs are exported from the nucleus to the cytoplasm with the help of exportin 5 in a Ran-GTP-dependent manner, containing a stem loop, a phosphate group at the 5' end and 2-nt 3' overhang. Then, the pre-miRNA is recognised and further processed to an miRNA duplex by an endonuclease complex comprising the RNase-III-like enzyme DICER, the transactivating response RNA binding protein (TRBP) co-factor and protein kinase R-activating protein (PACT). The mature miRNA is a ~22-nt double-stranded RNA molecule, with phosphate group at both 5' ends and 2-nt 3' end overhangs. The miRNA duplexes are then bound by

the RNA-induced silencing complex (RISC), which associates with one of the four Argonaute proteins (AGO1-4). miRNAs with central imperfect complementarity preferentially associate with AGO1, whereas those with perfect complementarity associate with AGO2, the only AGO protein with catalytic mRNA cleavage activity. RISC contains an ATPase/RNA helicase domain which unwinds the miRNA duplex into guide and passenger miRNA, with the former to be associated with the RISC and the latter to be degraded. Selection of which strand will become the guide miRNA is based on the relative stability of miRNA duplex termini.

Usually the activated RISC with guide miRNA targets mRNAs with imperfect complementarity in their 3' UTR, leading to translational repressions and/or mRNA destabilisation. In the case of near-perfect complementarily, silencing is achieved by mRNA–cleavage instead. The specificity of the miRNA is characterised by the seed sequence, which is situated 2–8 nucleotides from the 5' end of the miRNA, and the presence of target sequences in the 3' UTR [150].

Exploiting the above natural mechanism of gene expression regulation, silencing of genes can be achieved with the delivery of chemically synthesised exogenous siRNA that mimic the structure of the processed miRNA duplex (~22 nucleotides with 2-nt 3' overhang) but possess perfect complementarily with their intended target mRNA, therefore resulting in mRNA cleavage with the association of AGO2. Even though using siRNAs is a fast and efficient silencing approach, it has three major drawbacks: a) siRNAs have a short-half-life (susceptible to RNase), b) siRNAs require an efficient delivery system and c) their knock-down effect will be transient even for siRNA-derivatives with enhanced stability, since the concentration of the siRNA will drop with cellular divisions [150].

The above problems are overcome with the use of short-hairpin RNAs (shRNAs), which may be expressed from plasmid or viral vectors. In this case, shRNA can be continually expressed driven by an RNA pol3 promoters, such as U6 and H1, or in a more regulated and tissue-specific manner under an RNA pol2 promoter. Expression under the U6 promoter ensures high and continuous expression of the shRNAs, which may, however, result in cytotoxicity, increased off-target effects and the activation of immunological responses. Furthermore, components of the RNAi mechanism, such as exporting 5 and AGO2, can become saturated with overexpression of shRNAs, great abundance of which can therefore interfere with physiological function of the endogenously expressed miRNAs [150].

Further optimisation of RNAi strategies uses combinatorial approaches between shRNA and miRNA, which includes the insertion of siRNA sequences within the sequence framework of a naturally occurring miRNA at the locations of the original miRNA hairpin sequence. These constructs are known as shRNA-miR constructs and seem to reduce the toxicity and cell-death caused by interfering with the endogenous miRNA pathways and by triggering the interferon response [150–153].

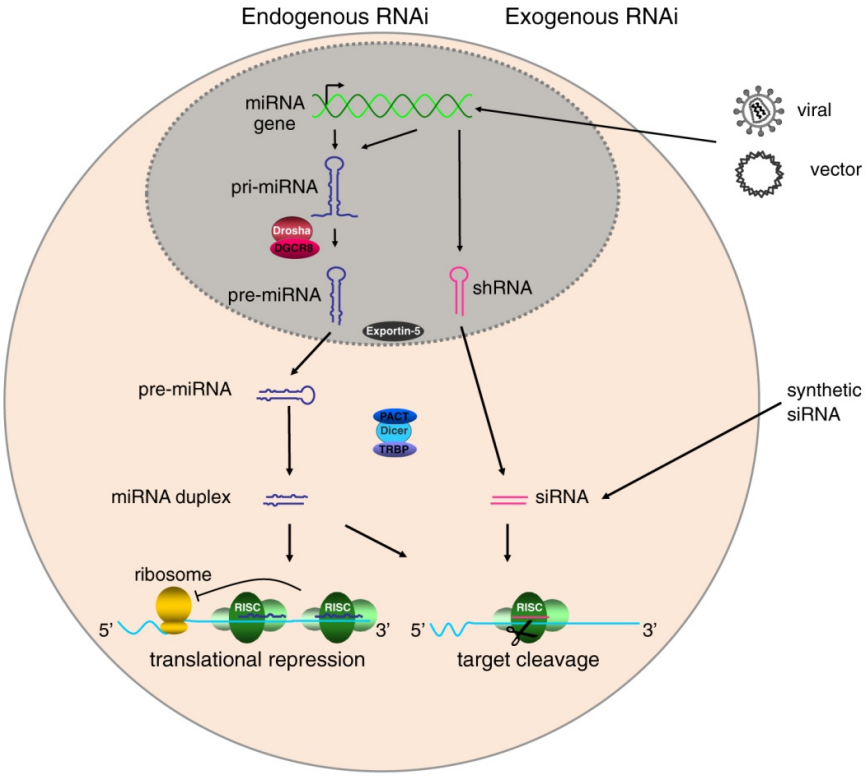


Figure 13: RNA interference (RNAi) pathway.
RNAi can be induced either endogenously or exogenously. The endogenous RNAi pathway starts with the expression miRNA genes for the production of long pri-miRNAs transcripts which form distinctive imperfect hairpin structures. These are then recognised and processed by the nuclear RNase-III-like endonuclease Drosha and its co-factor DGGR8 to ~70-nucleotide dsRNA molecules known as pre-miRNAs. Pre-miRNAs are exported to the cytoplasm by Exportin-5 and further processed to a ~22 bp imperfect miRNA duplex by the DICER/TRBP/PACT endonuclease complex. The one strand of the miRNA, aka guide miRNA, directs the RISC complex to the target mRNA for silencing. Imperfect complementarily of miRNA with the target mRNA leads to translational repression, whereas perfect complementarily, which is also achieved with siRNAs, results in mRNA cleavage. Exogenous RNAi can be achieved with the expression of shRNAs constructs either after plasmid DNA transfections or viral vector delivery. shRNAs are processed by DICER/TRBP/PACT endonuclease complex into siRNA with perfect complementarily. Exogenous RNAi can also be achieved with the direct transfection of synthetic siRNAs. Finally, shRNA-miR (not shown) can also be delivered to the cells as plasmid DNA or viral vectors, and shRNA-miR transcripts follow the pre-miRNA processing pathway. In this case prefect complementarily with the target mRNAs leads to mRNA cleavage. shRNA, short hairpin RNA; miRNA, microRNA; pri-miRNA, primary microRNAs; DICER, RNase-III-like enzyme; TRBP, transactivating response RNA binding protein; PACT, protein kinase R-activating protein; RISC, RNA-induced silencing complex. Reproduced from [150]

1.9.4.2 *RNAi therapeutics in β-haemoglobinopathies*

RNAi technologies have been applied for the experimental treatment of many diseases, including β-thalassaemia. A promising RNAi target for the reduction of disease severity in β-thalassaemia is excess HBA chains, as detailed in section 1.7.3.4. After confirming the beneficial effect of coinheritance of α-thalassaemia on β-thalassaemia disease severity in their murine model [154], Voon *et al.,* attempted to correct the HBB/HBA chain imbalance by using siRNA against the HBA mRNA *in vitro* on differentiated β-thalassaemia murine primary erythroid cells [155]. Using this transient approach, RNAi of HBA achieved 50% knock-down of *HBA* expression compared to normal control levels and clear phenotypic correction [155]. In the same vein, Xie *et al* attempted to restore the

imbalance of HBB/HBA chain ratio by microinjection of LVs in single-cell humanized β^{654} (β^{+}, IVSII-654 missplice mutation, HGVS: HBB:c.316-197C>T) mouse embryos, encoding an asRNA for the restoration of splicing and/or an shRNA against HBA mRNA. All HBA-targeting animal models and their offspring indicated a reduction of HBA mRNAs levels by 20–35% and a sustained amelioration of haematologic parameters, ineffective erythropoiesis and EMH [156]. The same group delivered the antisense and siRNAs in plasmid vectors instead of LVs into the tail veins of the β654 (β^{+}, HGVS: HBB:c.316-197C>T) mouse model and achieved a ~50% knock-down of HBA mRNA, a reduction of poikilocytosis, target cells and reticulocyte counts, but without improvements to red cell counts and Hb levels [157].

Ten years ago, in an innovative approach combining gene addition with RNAi, Samakoglu *et al.* developed the G9 RNAi vector, which encodes a synthetic *HBG* gene that allows expression of anti-sickling shRNA ($shRNA^{SCD}$) from untranslated regions of the precursor RNA [158]. The recombinant *HBG* transgene for treatment of SCD was expressed under the same *HBB* promoter, enhancer and βLCR elements as described in May *et al.* for the TNS9 vector [126]. In addition, the second intron, 5' and 3' UTRs were also derived from the *HBB* gene and were used to insert the $shRNA^{SCD}$-encoding sequence in three different positions for independent vectors, two in intron 2 and one in the HBB 5' UTR, as shown in Figure 14. Of the three positions chosen for shRNA expression, the one close to the SA site of *HBB* intron 2 in the vector-encoded *HBG* gene showed the highest knock-down efficiency for the targeted sickling mRNA. In particular, β^{s} transcripts were reduced by 74% and 96% in β^{S}/β^{S} homozygote and β^{0}/β^{S} compound heterozygote patient cells, respectively, whereas the *HBG* expression was similar in samples transduced with or without the anti-sickling shRNA [158]. The approach demonstrated the feasibility of shRNA-mediated knock-down from any LV expressing a HBB-like globin, and thus suggested a means of targeting aberrant mRNA species for degradation using the GLOBE vector. Co-expression of the globin transgene and the anti-sickling siRNA from the same construct under an RNA pol2 promoter in an erythroid specific manner will potentially reduce the toxic side effect developed by the constitutive and high expression of siRNA in cells, such as off-targeting, perturbation of the normal miRNA pathways and induction of interferon response.

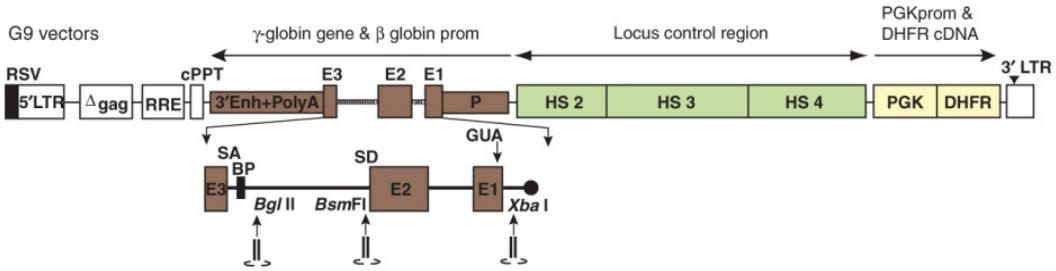


Figure 14: G9 HBG RNAi anti-sickling lentiviral vector.
G9 holds a recombinant *HBG* transgene with *HBB* IVSII. In addition, the transgene is driven under a *HBB* promoter (P), enhancer and βLCR elements described also previously by [126] for the TNS9 vector. The anti-sickling shRNAs sequence were cloned in three different positions: a) *Bgl*II restriction site (RS) located near the IVSII branch point (BP), b) *BsmF*I RS located near the IVSII splice donor site (SD) and c) *Xba*I RS in the HBB 5' UTR. PolyA, human *HBB* polyadenylation signal, 3'Enh, 3' HBB enhancer; E1–3, *HBG* exons; β-globin Prom, human *HBB* promoter; HS, Locus control region DNase I hypersensitive element; LTR, Long terminal repeats; Δgag, deleted gag region;, RRE, Rev responsive element; cPPT (central polypurine tract; PGKprom, phosphoglycerate kinase promoter; DHFR, mutant dihydrofolate reductase cDNA (Methotrexate resistance (MTXResistance); AUG, start codon.

Other groups aim to correct the imbalance of HBB/HBA-like globin chains in β-thalassaemia by activation the expression of the endogenous *HBG* genes, thereby increasing the amounts of HbF. As detailed in section 1.7.3.3, genes have been identified that are involved in the regulation of the foetal-to-adult globin switching and for which mutations are linked with the developments of HPFH disease. Factors, such KLF1, Oct1, Myb and BCL11A, which are responsible for or involved in *HBG* suppression in adults, constitute ideal candidate targets for RNAi, with several groups pursuing this approach [39]. An increase of HbF would also be a therapeutic for the treatment of SCD, owing to the anti-sickling properties of HBG chain [159], so that this approach is also of significant financial interest. Unsurprisingly, therefore, the most promising study to date on RNAi-based induction of HBG by Brendel *et al.,* was targeted on SCD [160]. Based on RNA pol2-based shRNA expression and the associated benefits (see above) they successfully developed a βLCR-shRNA^{miR} LV, in which a miRNA-adapted shRNA against BCL11A is expressed in an erythroid-specific manner under the control of an HBB- promoter and βLCR (HS2/HS3). Transduced normal or SCD HSCs exhibited 90% reduction of the BCL11A protein expression which led to an increase by 50%-60% of *HBG* expression and finally increase of HbF compared to the untransduced normal or SCD HSCs. In addition, they showed that lineage-specific silencing of the BCL11A overcomes the engraftment defect of HSC in mice observed when shRNA^{miR} was ubiquitously expressed under a spleen focus-forming virus (SFFV) promoter. In general the efficient increase of HbF in combination with the selective survival advantage of non-sickling RBCs suggests the application of the underlying tools for SCD GT clinical trials.

1.9.5 *Gene therapy by gene correction*

GT treatments using gene addition have by now proven successful in the preclinical treatment of countless monogenic diseases, are being tested in many clinical trials [161], have already achieved multiple orphan drug designations and have in several instances finally awoken commercial interest and achieved approval for drug marketing in the European Union [162]. Despite their promising and successful application, these treatments still have several inherent limitations in safety and efficiency, which have become apparent in part through the studies detailed above and which prompt a continued search for superior, next-generation treatments. In brief, one of the major concerns of permanent viral GT is the risk of insertional mutagenesis, which might lead to adverse events by deactivation of tumour-suppressor genes or activation of proto-oncogenes. Variable expression levels or long-term transcriptional silencing of integrated therapeutic genes are an additional and in part related concern that may be addressed by inclusion of barrier, enhancer or insulator elements in the provirus; an approach, however, that is limited by the packaging constraints of viral vectors and a negative impact of these elements on vector production and stability. Moreover, gene augmentation for some diseases may require the delivery of large transgenes, with the 70-exon, 3400-codon full-length dystrophin as an extreme example, which often exceed size constraints of suitable delivery vectors and in consequence require compromises between size and efficiency of the therapeutic transgene, with the choice of size and number of βLCR elements as a pertinent example. Gene addition is unsuitable for the treatment of disorders with gain-of-function causative mutations, for which concurrent RNAi knock-down of the aberrant gene product would be required. In turn, combinatorial approaches with RNAi or use of RNAi on its own do not provide full abolishment of the target gene expression and require tightly regulated expression in order to avoid or minimise off-targeting side effects, perturbation of normal miRNA pathways and activation of the interferon response by excess dsRNA. In addition, RNAi is limited to diseases in which gene target knock-down is beneficial and complete abolishment of the expression is not required to obtain a therapeutic effect [163].

The above limitations can be overcome by genome-editing approaches, which can achieve targeted and controlled modifications on the genome, providing stable long-term therapeutic benefit and allowing correction or primary defects at the source, for gain or loss of function mutations and viral infections alike. Gene correction potential was shown in earlier studies, which relied on homologous recombination (HR) events that occurred at very low frequencies in somatic cells (10^{-6}) [164,165]. A paradigm shift came about by the discovery by Mary Jasin that double strand breaks (DSB) on genomic DNA (gDNA) increase the frequency of genome editing by HR by 2–5 orders of magnitude,

providing overall frequencies of more than 5%, and formation of targeted insertions/deletions (INDELs) by nine orders of magnitude [166,167]. The induction of targeted DSB has since played an essential role in the development of more efficient genome editing, in combination with the discovery and invention of the corresponding programmable designer endonucleases. Of these, four platforms have emerged as those most frequently used for targeted DSB creation, specifically meganucleases (also known as homing nucleases), zing-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and, as the most recently developed and most easily applied editing platform, clustered regularly interspaced short palindromic repeat (CRISPR)-associated nuclease Cas9 [163].

1.9.5.1 *Repair pathways for genome editing*

The increased efficiencies of genome editing due to targeted DSB relies on endogenous DNA repair pathways. DSB repair can be performed by two major pathways, homology-directed repair (HDR) and non-homologous end joining (NHEJ) [168]. HDR is a rare and precise pathway that naturally uses the homologous sister chromatid as template for accurate correction. As a result, HDR is restricted to the late-S/G2 cell cycle phases, when DNA replication is completed and sister chromatids are available to serve as repair templates [169,170]. This pathway can be exploited for the correction of mutations or the insertion of therapeutic transgenes at a specific site, by providing as template double strand DNA (dsDNA) or a single strand DNA (ssDNA) with significant sequence overlap (e.g. >400-bp homology arms) with flanking regions of the target site. Transgene insertion may target safe harbour loci, such as the adeno-associated virus site 1 (AAVS1) and the human chemokine (C-C motif) receptor 5 (*CCR5*) gene locus in haematopoietic stem cells, driven by their own transgene promoters [171,172]. Another approach is the insertion of the therapeutic transgene as a promoter-less cDNA at the start site of endogene coding sequence, to place it under the control of the endogenous promoter. This will provide regulated expression of the cDNA and correct mutations downstream of the cDNA insertion site, while defects in promoter or upstream regulatory regions cannot be addressed plus alternative splicing and regulatory function of intronic sequences associated with the normal gene expression will not be reproduced (Figure 15) [172]. In contrast to HDR, the NHEJ repair pathway is fast (with 10–30 minutes operating half-times)[173]. NHEJ constitutes the prominent cellular DSB repair pathway, which is template independent, error-prone and is active in all phases of the cell cycle [168]. Most of the DSB repairs therefore occur through the NHEJ repair pathway, by which a direct re-ligation of the two DNA ends is completed in a non-mutagenic manner. However, repeated induction of targeted DSB eventually leads to the formation of INDELs at the DSB site [166,172]. Targeted-formation of INDELs via NHEJ is used to knock out genes and disrupt miRNA-encoding and

regulatory genetic elements. It has also been used to restore the correct reading frame caused by a disease-causing frameshift mutations or to disrupt normal or aberrant splicing sequences in order to achieve exon skipping or restoration of splicing [174–176]. When two simultaneous DSBs occur in *cis*, large defined deletions or, more rarely inversions (1% of deletions), can take place [166]. The former can be used to remove a pathogenic exon, a set of exons or harmful trinucleotide repeat expansions (such as those causing Huntington disease, myotonic dystrophy or Friedreich's ataxia) from a gene. Inversion may also serve the inactivation of a target gene, but may in principle be used to cure in human diseases brought about by chromosomal inversion, as is the case for haemophilia A caused by inversion within the *Factor VIII* gene [172]. Even though TALENs and CRISPR/Cas9 have become more popular than ZFNs in their recent research and preclinical application, mainly due to their easy accessibility, ZFNs are as yet the only editing tools to have entered clinical trials. This is the case for the targeting of the *CCR5* gene in autologous $CD4^+$ T cells from patients with HIV (clinical trial number NCT00842634), where disruption of the CCR5 receptor for HIV attachment is required for successful infection and where edited cells resulted in reduced levels of HIV in blood in the majority of the patients after reinfusion [177].

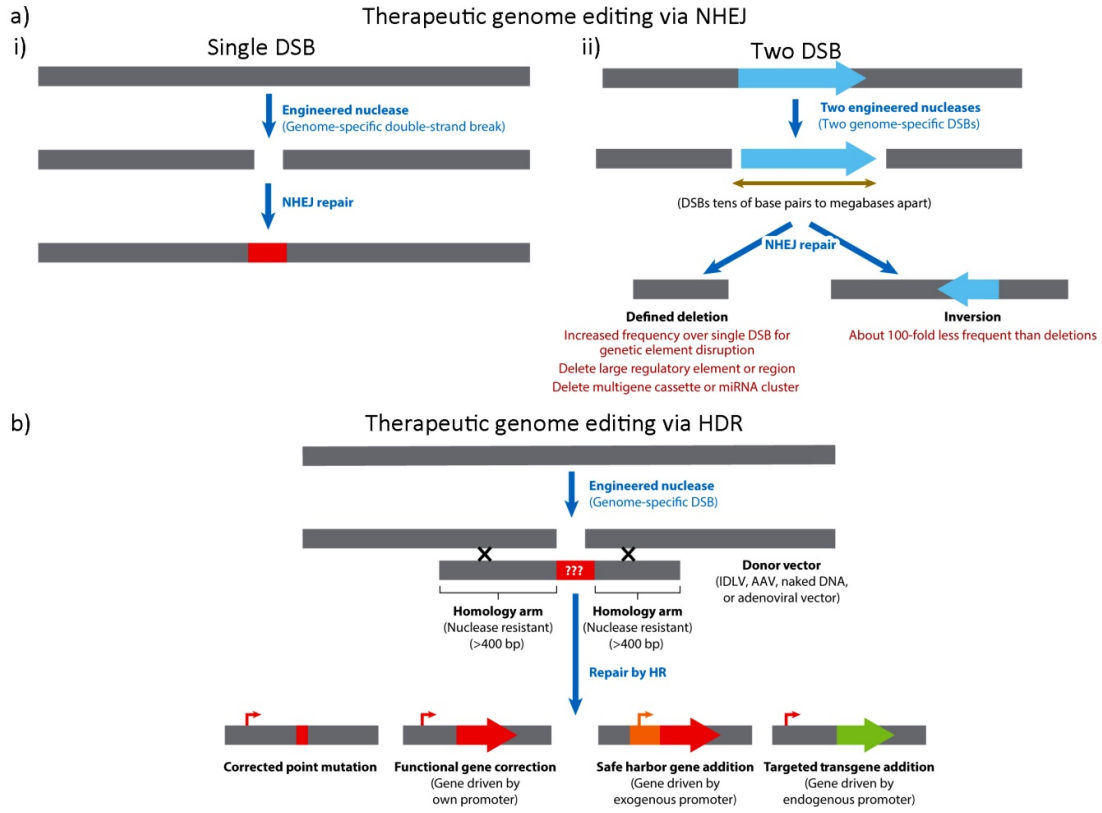


Figure 15: Therapeutic genome editing.
The overview shows a) non-homologous end joining (NHEJ) of i) single or ii) two simultaneous double strand breaks (DSB) and b) homology directed repair (HDR) as major pathways of DSB repair. IDLV integrase-deficient lentiviral vector, AAV adeno-associated viral vector. Modified from [166]

1.9.5.2 *Meganucleases*

Meganucleases are produced by re-engineering the DNA-binding domain of naturally occurring homing endonucleases, such as the most commonly used I-*Sce*I and I-*Cre*I, in order to target novels sequences. They have large DNA binding sites (>14 bp), and their DNA binding domain also performs DNA cleavage [163]. Meganucleases are the smallest type of engineered nucleases, which allows their delivery with all standard gene delivery methods and which allows multiplexed genome editing by packaging of multiple meganucleases in a single virion [178]. Upon DSB-formation these nucleases lead to the formation of 3' overhangs which may be more recombinogenic for HDR than the mainly 5'overhangs which are produced the *Fok*I endonuclease used in combination with ZFNs and TALENs [178]. The major impediment to wider application of meganucleases in genome-editing approaches is the difficulty and laboriousness of re–engineering their specificity for the cleavage of new target sites.

1.9.5.3 *Zinc-finger nucleases*

Zinc-finger nucleases are artificial proteins in which zinc-finger (ZF) DNA binding domains are fused with the non-specific cleavage domain of *Fok*I endonuclease, which belongs to the IIS restriction enzymes. Since *Fok*I acts as dimer, ZFNs act as pair, where two monomers need to bind in close proximity (with a 5- or 6-bp spacer) and in the right orientation in order to be active. Each ZF consists of approximately 30 amino acid and forms a ββα structure coordinated by a zinc ion, with residues in the α-helix recognising nucleotide triplets in the major groove of the DNA and with the affinity and specificity of one ZF domain affecting those of the adjacent ZFs [178–180]. In general, DNA binding domains of ZFNs range between 9–18 bp, which confers high specificity to their target sequence, with low off-targeting potential. ZFN DSB-formation produces defined 4- or 5-bp 5' overhang [181]. Even though these tools have proven highly efficient in genome editing, with one example having entered clinical trials (CCR5 ZFN) [177], designing long-arrays of ZFs in order to achieve unique binding specificity in the genome despite context-dependent interaction of neighbouring ZFs constitutes a major bottleneck in their application to new targets and thus limits their broader use for genome editing. ZFN are small enough for the packaging of both monomers into a single virion for co-delivery to cells as pairs. Moreover, since ZFN act as pairs, their combined recognition sequence confers a lower risk of off-target cleavage, which is a major concern particularly in the clinical application of genome editing, relative to monomeric designer nucleases, such as meganucleases and standard CRISPR/Cas9 [179].

1.9.5.4 *Transcription activator like effector nucleases*

TALENs are artificial proteins in which TALE DNA binding domain are fused with the non-specific cleavage domain of *Fok*I endonuclease. TALEs are virulence factors found in the plant pathogenic bacteria genus *Xanthomonas.* The DNA binding domain comprises a range of conserved repeats of 33–35 amino acids, with each repeat responsible for the recognition of a single nucleotide [179]. The nucleotide specificity is determined by two hyper-variable amino acids at positions 12th and 13th (repeat-variable di-residues (RVDs) of the conserved amino acid sequence in each repeat. In consequence, the choice of RVDs determines the DNA recognition sequence in a clear amino-acid code, NN, NG, NI, and HD preferentially recognising guanine, thymine (T), adenine (A), and cytosine (C), respectively. The elucidation of the TALE RVD code, in which specific RVDs were identified to interact and bind to specific nucleotides was a watershed for the use of TALENs for genome editing [182,183]. Customised TALENs have long DNA binding domains of 13–28 repeats, which in combination with their mandatory action as dimers makes them highly specific designer nucleases with low off-targeting potential [184]. The major advantage over meganucleases and ZFN is the ease of designing new TALENs, which has prompted their use by many groups. The only main requirements for TALEN design are the presence of a thymine (T) in the 5' end of the target sequence and a spacer between the two TALEN monomers of between approximately 10 and 15 bp [185,186]. Different platforms have been developed for engineering TALENs, the most accessible systems allowing assembly of new TALENs by standard cloning techniques and the use of plasmid libraries for which each construct encodes a single TALE repeat. Medium-throughput assembly methods were developed that rely on the Golden Gate cloning method and allow assembly of TALE multimers in single-tube reactions [187,188]. High-throughput assembly platforms have also been developed, such as the solid-phase assembly and other systems based on ligation-independent cloning techniques [189]. TALENs mainly create 5' overhangs, but may also create 3' overhangs, and the majority of the INDEL modifications are deletions, while ZFNs produce a balanced distribution of insertions and deletions [181,190]. The large size of TALENs (~4 kb) and the repetitiveness of their sequence constitute a major drawback for their delivery to cells. First, delivery of both TALEN monomers in a single viral vector with limited package capacity is a challenge if not impossible for many viral platforms. Second, repetitiveness of the conserved parts of their coding sequence impairs packaging in viral vector, owing to the frequent occurrence of re-arrangements by recombination, as shown with the delivery of the TALENs using LVs [191]. However, for LVs this can be minimised by codon diversification between the repeats, which removes sequence identity at the nucleotide level [192], and moreover appears not to affect delivery of TALEN monomers by adenoviral vectors [191].

1.9.5.5 *Clustered regularly interspaced short palindromic repeat-associated nuclease Cas9*

Recently novel programmable RNA-guided DNA endonucleases (RGEN) have been discovered, as a promising alternative to other programmable endonucleases that allows design of target specificities based on Watson-Crick base-pairing between RGEN RNA and the target DNA. The clustered regularly interspaced short palindromic repeat-associated nuclease Cas9 system is part of the bacterial adaptive type-II immune system against foreign DNA through RNA-guided cleavage, and consist of two RNA components, the CRISPR RNA (crRNA) and the trans-activating RNA (tracrRNA), and one protein component, the CRISPR-associate protein (Cas9) [193]. The natural immune response is split into two events; the memorisation of the invading foreign element (adaptation) and the destruction of the element (interference). The adaptation event occurs upon first infection, when Cas proteins recognise the foreign elements and insert DNA fragments of that element in the CRIPSR array in bacterial genome, which constitutes the genetic memory that efficiently prevents the re-infection of the bacterium with the same element [193]. The CRISPR array is transcribed and undergoes one or two processing events leading to the production of many mature crRNAs. crRNA/tracrRNA duplexes bind Cas9 to form an active CRISRP ribonucleoprotein complex, ready for sequence-specific degradation of cognate invading DNA [194].

In 2013, two studies showed how to successfully engineer the type-II CRISPR system from *Streptococcus thermophilus* and *Streptococcus pyogenes* in order to achieve NHEJ- or HDR-mediated genome editing in mammalian cells. This can be done by heterologous expression of the tracrRNA and crRNA or, as described by Doudna, Charpentier and colleagues in 2012 for the first time, by expression of a hybrid single guide RNA (gRNA), which encodes the customised crRNA target sequence in the framework of the tracrRNA sequence [195–197]. As in the case of crRNA, the gRNA contains a 20-nt guide sequence which is responsible for CRISPR/Cas9 specificity to the target site. The simplicity of the design and development of different CRISPR/Cas9, as also the possibility of targeting multiple genes at once with the delivery of multiple gRNAs are some of the main reasons this platform has become the most popular genome-editing tool for a range of biological applications [194]. As one key limitation of the system, completion of DNA cleavage by Cas9 (via the HNH and RuvC nucleases domain for the sense and antisense strand, respectively) after target binding depends on the presence of a protospacer-adjacent motif (PAM) sequence immediately downstream of the target site. Only when Cas9 and sgRNA successfully bind to both PAM and target sequence is Cas9 activated and cleaves the target DNA. Each Cas9 protein has a specific PAM sequence, which in the case of the widely used *Streptococcus pyogenes* Cas9 is NGG, with less efficient cleavage also for sequences carrying the sequence NAG instead. [194,198,199]. The small size of

sgRNA and Cas9 cassettes allows the efficient delivery of both constructs in a single viral vector such as the adeno-associated virus (AAV). Furthermore, in contrast to the TALENs, CRISPR/Cas9 cleavage efficiency is not affected by DNA methylation [198]. Due to the large diversity of the natural CRISPR system, it is expected that many novel CRISPR-based gene-editing technologies will be discovered in the near future, such as the already described RNA-guided endonucleases Cpf1 [200].

1.9.5.6 *Optimisation of designer nucleases specificity*

One of the major concerns for the use of programmable nucleases is their off-target cleavage potential, in particular for CRISPR/Cas9, which acts as monomer and therefore poses a greater the risk than ZFN and TALENs. Increased off-targeting is associated with increased toxicity in edited cell populations [190] and moreover raises the same concerns over potential adverse events as the potential of insertional mutagenesis for gene-addition approach. CRISPR/Cas9 specificity can be increased with the use of modified Cas9 nickases in which the HNH or RuvC nuclease domain is inactivated, requiring dimeric action in order to achieve a DSB [201]. In an attempt to adopt the benefits of *Fok*I dimerisation activity, a hybrid inactivated Cas9 protein fused with the *Fok*I cleavage domain has been developed, Cas9-Fn, which has higher specificity than the wild-type Cas9 [202]. In addition, reducing the length of complementarity with the target sequence from 20 bp to 17 bp rather counter-intuitively increases the cleavage specificity of spCas9 [203]. For *Fok*I-containing designer nucleases, a dramatic increase in on-target specificity is achieved with the modification of the *Fok*I dimerization domain to force ZFN and TALENs (ELD/KKR) *Fok*I heterodimer, so that in order for the nuclease to be active, binding of both different target sequences with the right orientation and spacing is required [204,205]. Finally, delivery of the designer nuclease as mRNA or protein to the cells reduces off-targeting due to their shorter half-life relative to the plasmid DNA delivery, which minimises their exposure to gDNA and potential off-target sites. In line with this, off-targeting is directly dose dependent [178].

1.9.5.7 *Application Emphasis on Thalassaemia*

Targeted gene correction of β-thalassaemia mutations provides a safe and permanent treatment without the risk of insertional mutagenesis and with *per definitionem* normal expression levels of the corrected endogenous *HBB* gene [66]. For gene correction, artificial nucleases with customisable DNA binding domains are used to create site-specific DSB, which can then be repaired through HDR, using as template homologous donor DNA with the desired sequence change, or by NHEJ, which disrupts a sequence by INDELs at the site of the DSB [166]. In the specific instance of IVSI-110, for example, both HDR-mediated repair of the mutation and NHEJ-mediated disruption of the aberrant splice site

would be repair options, with greater precision of the former and higher efficiency of the latter. Gene correction is the definitive personalised medicine approach. More universally while less efficient in the individual, the same tools could be applied to activate endogenous *HBG*, e.g. by disabling *HBG* repressors or downstream promoters on the *HBB* locus. It has recently been shown in hCD34^{+} cells by disrupting the +58 DNase hypersensitive site (HS) erythroid-specific enhancer element of BCL11A using different designer nucleases, that upregulation of *HBG* may be achieved comparable to that seen upon disruption of the *BCL11A* coding sequence [206–208].

Accurate gene correction of mutations through the HDR pathway and the delivery of exogenous DNA templates are hindered by the relatively low frequencies of HDR compared to NHEJ. However the therapeutic potential of accurate gene correction in cells may still be exploited for β-thalassaemia, regardless of potentially low efficiencies, since based on previous allogeneic BMTs a mixed BM chimerism of 20% is sufficient for transfusion independence. This is due to the survival advantage of healthy/corrected compared to thalassaemic erythrocytes. [208,209] In addition, Voit *et al.* demonstrated a different approach for *HBB* gene correction by TALEN-mediated targeted integration of the HBB cDNA with approximately 20% efficiency at the endogenous *HBB* start codon in K562 erythroleukaemia cells [210]. Gene correction of the *HBB* gene was also achieved in induced pluripotent stem cells (iPSCs) for β-thalassaemia and SCD. In particular, Ma *et al.* corrected the mutation in iPSCs from thalassaemic patients compound heterozygous for the HBB$^{IVSII-654}$ (β^{+}, HGVS: HBB:c.316-197C>T) and β41-42 (del TCTT) (β^{0}, HGVS: HBB:c.124_127delTTCT) mutations, using specific TALEN pairs and a linearised plasmid that encoded removable antibiotic selection markers flanked by LoxP sites and the whole *HBB* as donor template [211]. Others following the same strategy for the correction of iPSCs from SCD patients, achieved 30–40% expression of normal HBB transcript relative to uncorrected cells, after haploid repair, antibiotic selection of positive clones and erythroid differentiation [212,213]. A similar approach was followed by Xie *et al.*, using CRISPR/Cas9 to correct -28 A/G (β^{+}, HGVS: HBB:c.-78A>G) and β41-42 (del TCTT), Xu *et al.*, using both, TALENs and CRISPR, to correct HBB$^{IVSII-645}$, and by Sun and Zhao, using TALENs for the correction of SCD. However, in these studies, piggyBac transposons were used to allow seamless removal of the selection gene and in order to improve upon Cre-Lox-based removal of resistances, which leaves a footprint of the ectopic sequence at the site of correction [214–216]. Most importantly for these studies, the successful correction of patient-derived iPSCs was achieved, maintaining their pluripotency, preserving normal karyotype and allowing efficient differentiation into haematopoietic progenitors without generation of off-target effects. However, iPSCs are still not ready for clinical applications because of concerns over their long-term safety, which requires further assessment before their use in human subjects. Therefore, genome editing of autologous CD34^{+} cells is still the gold standard for clinical application

of GT approaches, vindicated by the successes of current clinical trial for β-thalassaemia and SCD by gene-addition GT [131]. In the case of gene correction of HSCs, clinical application is still hindered by low efficiencies of the approach in the absence of clonal or enrichment steps (e.g. by cell sorting) [39]. However, in a recent study 10–20% gene correction of SCD mutation was achieved for hCD34^{+} cells using ZFNs mRNA and delivery of donor DNA template by an integrase-deficient lentiviral vector (IDLV) or as short oligonucleotide DNA. Subsequent transplantation of the corrected cells into NOD/SCID/IL2rγ^{null} mice gave markedly reduced gene correction levels 16 weeks post-engraftment in spleen and BM, [217], suggesting that HDR mediated-genome editing is not efficient in the long-term HSC subpopulation. This finding is in agreement with previous work by Genovese *et al.*, who achieved significant HDR-mediated correction of X-linked severe combined immunodeficiency also in LT-HSCs by improving tolerability of LV treatment through modified transduction and culture conditions [217,218]. In a more recent study, Hoban *et al.* achieved correction of 20 % of the hCD34^{+} from SCD using CRISPR/Cas9, by delivery of Cas9 mRNA by electroporation and of gRNA and donor template with IDLV [219]. While these efficiencies move HDR-based gene correction into the realm of clinical applicability, NHEJ approaches by far exceed these efficiencies and are the repair pathway of choice for application in HSC, where the therapeutic target allows [208,220].

1.10 Research project: Aims and Objectives

β-Thalassaemia, as detailed in section 1.7.3, is caused by more than 400 mutations affecting the synthesis of functional HBB chains at different stages of expression : at the transcriptional, posttranscriptional RNA processing or translation stage [56]. The type of the mutation (β^{0} ,β^{+} and β^{++}) in combination with disease modifiers (see section 1.7.3.3) is responsible for the wide range of β-thalassaemia phenotypes (major, intermedia and minor) for which each thalassaemia patient ought to receive specific treatment in order to warrant the best possible outcome, be it of disease management or of curative therapy, such as GT [56]. For example, a GT protocol optimised in order to reach therapeutic levels of *HBB* expression might differ between homozygote patients for the β^{0} cd39 mutation and those with β^{+}IVSI-6 or IVSI-110, as the latter show residual HBB expression and thus potentially lower requirements for levels of gene expression per vector copy, for the number of vector copies per cell required or for the harshness of conditioning regimens. This was also indicated by Breda *et al.* where differentiation of erythroid progenitor cells from β^{0}/ β^{+} and β^{+}/ β^{+} patients transduced with AnkT9W LV (VCN 0.6 ≤), a *HBB*-expressing LV with an erythroid-specific ankyrin 5' HS insulator [221], lead to the production of total HbA at levels comparable to those of control cells from disease carriers, whereas in the case of β^{0}/β^{0} cells, higher VCNs were required [114]. Therefore, it is likely to achieve transgene therapeutic levels at lower VCN with patients β^{+}/β^{+} or β^{+}/β^{0} than with

β^0/β^0, and if this observation is to be taken into account during treatment, it will help minimise the risks of insertional mutagenesis or avoid full chemical myeloablation [114]. Development of personalised GT could further optimise the safety and efficiency of GT tools, especially in patients where the efficiency of gene-therapeutic correction of disease parameters in not increased in a VCN-dependent manner, thus indicating the relevance of additional, patient-specific factors [114,222].

The aim of my PhD project is the development of a personalized GT for patients with specific mutations that lead to the production of persistent, non-functional mRNA, such as the common *HBB*$^{IVSI-110(G>A)}$ (IVSI-110) mutation, following two different approaches, a) via RNAi-based gene silencing, possible in combination with gene addition and b) via genome editing. Necessary for both approaches is the development of IVSI-110 model cell lines, for the establishment of a proof of concept of our hypothesis and the shortlisting of candidate therapy tools before moving to more laborious and costly analyses in patient-derived $CD34^+$ cells. Therefore, we need to produce transgenic murine erythroleukaemia cell lines (MEL) holding the mutant version of the *HBB* expressing MA821 vector (MEL_MA821_HBB^{normal}) and its mutant derivative (MEL_MA821_HBB$^{IVSI-110}$) at average VCN ~2 and clones with defined VCNs, e.g. VCN 1, after two rounds of clonal selection. These lines will allow the analysis of RNAi or genome-editing action without additional endogenous targets and without a background expression of human *HBB*, with the matching MA821 line allowing the comparison of expression levels achieved to those approximately expected from the normal human *HBB* locus. Hand in hand with our mRNA-species-specific therapy approach we also need to develop an accurate and sensitive quantification method (e.g. multiplex reverse transcription quantitative polymerase chain reaction (RT-qPCR)) for the discrimination and quantification of the aberrant and normal HBB mRNA variants. Using the novel and already established quantitative methods, functional characterisation of the transgenic cells lines at a transcriptional and translational level needs to be assessed and correlated with the expression patterns seen in IVSI-110 thalassaemic patients.

Preliminary data from Breda *et al.* demonstrated that even though the aberrant mRNA from IVSI-110 patients is targeted by NMD pathway, it still constitutes 49% of the total HBB transcript population [222,223]. While IVSI-110 transcripts are NMD targets, their abundance might exceed a performance threshold of NMD surveillance, which requires a first round of translation for the recognition of PTCs [60,86,224]. The above study by Breda *et al.* also showed that in contrast to the β^0/β^0, erythroid progenitor cells from IVSI-110 β^+/β^+ and β^+/β^0 patients did not exhibit vector-dependent increases of HbA levels after T9W LV transduction, even though high transcription levels of the *HBB* transgene were detected [222,223]. This led them to the hypothesis that IVSI-110 mRNA is stable enough to

compete at a translational level with the expression of the therapeutic vector-encoded HBB protein, leading to the suggestion of a "traffic jam" model , in which IVSI-110 transcripts block ribosomes and impair the translation and therapeutic effect of vector-encoded *HBB* in IVSI-110 cells compared to β^0 (cd39) cells [222,223]. In line with this, Breda *et al.* also highlighted the need of careful pre-clinical investigation of the genetic background of the patients and the response of their cells to the therapeutic vector *in vitro* before their enrolment as study subjects in corresponding clinical trials [114,222,223,225].

Our RNAi- and gene-addition-based gene-therapy strategy aims to investigate whether degradation of aberrantly-spliced IVSI-110 mRNAs before their recruitment of translation factors and ribosomes might increase the translation of endogenous and vector-derived HBB mRNA to therapeutic levels. Our final therapeutic model includes the construction of a modified version of the GLOBE vector, MA821-T87Q, in analogy to the approach taken by Samakoglu *et al.*, where IVSI-110-specific shRNAs will be co-expressed with the vector-derived human T87Q-HBB from within the same transcription unit, to silence aberrantly spliced IVSI-110 mRNA variants [130,132,137,158]. The erythroid-specific expression of all shRNAs from a size-reduced version of the native LCR-*HBB* promoter will minimize their off-target effects. However, before testing the design of transgene and shRNA co-expression, we seek to establish a proof of principle of our hypothesis on the transgenic MEL MA821_HBB$^{IVSI-110}$ clonal cell line mentioned above, initially by delivering only IVSI-110-specific shRNAs with pLKO.I-based LVs, in which shRNA expression is driven by the constitutive RNA pol3 U6 promoter. Subsequently to the establishment of the proof of principle on MEL MA821 HBB$^{IVSI-110}$ transgenic cells, we will evaluate the therapeutic advantage of IVSI-110-targetting shRNAs for the treatment of IVSI-110 patient-derived hCD34$^+$ cells, alone or in combination with the MA821 (GLOBE) LV [132]. Validation of our hypothesis could have an essential impact on the development of optimised personalised GT, with numerous potential applications for similar β-thalassaemia mutations that introduce splice defects. RNAi alone or combination of RNAi and gene-addition approaches could enhance the therapeutic potential of the available *HBB*-like LVs, in order to allow patients to reach transfusion independence at lower VCN or for less efficient gene-addition vectors.

Our strategy of functional correction of causative *HBB* mutations via genome editing constitutes an attractive alternative to gene augmentation for personalised GT. As IVSI-110 creates an aberrant SA site, leading to the production of an aberrant non-functional HBB mRNA in consequence, our goal is the restoration of correct splicing via genome editing by disrupting the mutated nucleotide using state-of-the-art programmable nucleases (TALENs and RGENs) and exploiting the common and efficient NHEJ repair pathway. We believe that disruption of the aberrant SA site and/or the

adjacent sequences will shift pre-mRNA processing by the spliceosome towards the use of the normal SA site and lead to splice restoration. In order to achieve a targeted disruption of the IVSI-110 mutation we will design a site-specific RGEN and different sets of TALEN pairs (with varying specificity and affinity for their target sequence) to create a DSB at the site of or near the mutation. Initially, targeted disruption efficiency of all of our designer nucleases will be assessed in HEK293T cells. Then, as for our first approach, the most efficient therapy tools, in this case RGEN and TALEN pairs, will be used for the establishment of our proof of principle using the transgenic MEL MA821 $HBB^{IVSI-110}$ cell lines, where evaluation of splice restoration on TALEN- and RGEN-edited bulk populations will be assessed relative to both, the non-edited cells and the MEL_MA821_HBB^{Normal} counterpart transgenic cell line. In addition, we will perform full characterisation of the TALEN- and RGEN-induced INDELs and correlate the type of INDEL with functional data for splice restoration in edited clones. Subsequent to the evaluation of restoration of splicing on cell lines, the most promising designer nucleases will be tested in IVSI-110-patient-derived $hCD34^+$ cells, where targeted disruption efficiency and restoration of splicing will be measured relative to non-edited samples.

Our genome editing approach of functional correction of splice-site mutations by the NHEJ pathway will provide a more stable and effective correction of β-thalassaemia, compared to gene-augmentation approaches, since (for designer nucleases with optimised target specificity) insertional mutagenesis is avoided and HBB expression is still under regulatory control of the endogenous *HBB* promoter. In addition, our strategy is also superior in terms of efficiency and safety than the gene correction approaches followed by the majority of other groups, who mostly pursue accurate gene correction via the less efficient HDR pathway and who thus require iPSC clonal selection. Exceptional strategies based on the NHEJ pathway chosen by us were also pursued in parallel by Canver *et al.* and Traxler *et al.*, in order to achieve erythroid-specific suppression of the γ-globin repressor BCL11A in the first case and the emulation of a naturally occurring HPFH mutation on the HBB locus in the second[206,220]. Importantly, our therapeutic approach targets the authentic substrate for clinical gene-therapy of β-thalassaemia and SCD [39,131], $hCD34^+$ cells, and thus avoids persistent uncertainties over the clinical suitability of iPSCs. Finally, correction by the disruption of regulatory regions outside ORFs is not restricted to β-thalassaemia mutations. On the contrary, there are mutations suitable for this approach in over 180 genes, responsible for many human diseases in which the same approach could therefore have essential therapeutic effects ([81] and our own unpublished review of literature and mutation databases).

2 Materials and Methods

2.1 Standard molecular-biology methods

All molecular-biology techniques were performed based on kit manufacturer's instructions or according to Sambrook *et al.* 2001 [226]. These techniques include gDNA extraction and purification from eukaryotic cells *(Flexi Gene DNA Kit and Mini Blood DNA extraction kit; Qiagen, Hilden, Germany)* quantification of DNA quantity and quality by spectrophotometry *(ND-1000, Thermo Fisher Scientific Inc., MA, USA)*, enzymatic reactions such as restriction enzyme digest, polymerase chain reaction (PCR), ligations (*all New England Biolabs, NEB,* Frankfurt am Main*, Germany*), preparation and transformation of DH5α™ and One Shot® TOP10 chemically competent *E coli* cells (*Invitrogen™, Thermo Fisher Scientific*), Turbo chemically competent cells (*NEB*) and XL10-Gold ultra-competent cells (*Agilent Technologies, Santa Clara, USA*), bacterial culture, column-based plasmid DNA purification from bacteria (*NucleoBond Xtra Midi/Maxi kit, Macherey-Nagel GmbH, Duren, Germany*), agarose gel electrophoresis (using *NEB markers*) and gel-extraction of DNA fragments (*NucleoSpin® Extract II Kit, Macherey-Nagel GmbH, Duren, Germany*).

2.2 Agarose gel electrophoresis of DNA

Agarose gels where prepared at 0.8 – 2.5% concentrations, by dissolving the appropriate amount of Agarose powder (*Sigma–Aldrich, St Louis, USA*) in 100 mL of 1 x Tris Borate-EDTA (TBE) buffer and 0.2–0.5 μg/μL ethidium bromide (EtBr). DNA samples were mixed with loading dye (Bromophenol blue/Xylene cyanol FF or Orange G (*Sigma–Aldrich, St Louis, USA*) at 1x final concentration and loaded onto the gel along with 1 kb or 2 log DNA ladder (*NEB®, Germany*). Gels were run at 90 – 120 V/cm until the Bromophenol blue dye had migrated to approximately ¾ of the gel length. In general Bromophenol blue, Xylene cyanol and Orange G run at rates corresponding to 300, 4000 and 50 bp respectively on 1% TBE agarose gels. DNA fragments are then visualised using a UV transilluminator (*Vilber Lourmat, Wembley, Australia*) and images captured using the Mega-Capt software.

2.3 Genomic DNA extraction

2.3.1 *Genomic DNA extraction using Flexi Gene*

Genomic DNA extraction from transduced or genome edited MEL cells was performed using FlexiGene DNA kit (*Qiagen*), according to manufacturer's instructions. Briefly, 1–2x10^6 cells were suspended in 300 μL FG1 lysis buffer. Then 300 μL of FG2 denaturation buffer supplemented with 3 μL Qiagen Protease, mix with three time inversions, and incubated for 10min at 65 °C. Genomic DNA

was precipitated with 600 µL isopropanol followed by a centrifugation at 10 000 x g for 10 minutes. The precipitated gDNA was washed with 600 µL 75% ethanol (EtOH) and collected by centrifugation at 10 000 x g for 3 minutes. Genomic DNA pellet was left to air dry for 5–10 minutes before dissolving it in 200 µL of FG3 hydration buffer.

2.3.2 *Genomic DNA extraction using QIAmp DNA Mini Blood kit*

Genomic DNA from genome edited cells destined for T7E1 assay was extracted using QIAmp DNA mini blood kit (*Qiagen*), according to manufacturer's instructions. Briefly, 0.5–5 x 10^6 cells were suspended in 200 µL PBS supplemented with 20 µL Proteinase K and 4 µL RNase A (100 mg/mL). Then, 200 µL of AL lysis buffer was added to the samples, mixed by pulse-vortexing for 15 seconds. Samples were incubated at 56 °C for 10 minutes. At that point, 200 µL of EtOH was added to the samples, then the mixture was loaded to the QIAamp Mini spin column and centrifuged for 1 minute at 6 000 x g. Columns were washed with 500 µL of AW1 buffer and centrifuged for 1 minute at 6 000 x g and with 500 µL of AW2 buffer and centrifuged for 3 minutes at 20 000 x g. Residuals of AW2 buffer were removed with a centrifugation at 20 000 x g for 1 minute. Elution of the gDNA was achieved by loading 50–100 µL water to the columns, 1 minute incubation at room temperature (RTemp) and final centrifugation at 6 000 x g for 1 minute.

2.4 Plasmid DNA extraction

2.4.1 *Mini preps (Isopropanol / EtOH purification protocol)*

Bacterial pellets were collected by centrifugation of a 1 mL aliquot of culture at 17 600 x g (maximum speed) for 30 seconds in a microcentrifuge. Pellet was re-suspended in 250 µL re-suspension buffer 1 (50 mM Tris, 10 mM EDTA). With the addition of each buffer, mix of samples was by pipetting. Then, 250 µL lysis buffer 2 (0.2M NaOH, 1% SDS) and 250 µL of 3 M KAc/ 5 M Acetic Acid was added. The resulting lysates were centrifuged at 17 600 x g (maximum speed) for 10 minutes and supernatant transferred in new 1.5 mL microfuge tubes. Plasmid DNA was precipitated with 0.7 volumes of isopropanol and pelleted by centrifugation at maximum speed for 10 minutes, washed with 300 µL 70% EtOH and re-pelleted at maximum speed centrifugation for 5 minutes. The plasmid DNA pellet was left to air-dry for 10 minutes and then dissolved in 40 µL TE (25 mM Tris, 2.5 mM EDTA) buffer supplemented with 40 µg/µL RNase.

2.4.2 *Silica-based plasmid purification Midi, Maxi or Maxi EF Macherey Nagel*

Plasmid DNA extraction from midi and maxi bacterial cultures was performed using the Macherey Nagel NucleoBond Midi/Maxi kit (*Macherey-Nagel GmbH*) according to manufacturer's instructions.

Briefly, bacterial pellets from 250 / 500 mL cultures were re-suspended in re-suspension buffer, supplemented with RNase A, and lysed by addition of lysis buffer, mixed gently by five-times inversion and incubated at RTemp for 5 minutes. Then, neutralization buffer was added to the lysates and again mixed gently by inversion. Lysates were then loaded onto positively changed silica columns, to which plasmid DNA binds. The endotoxin-free maxi kit had extra washes of the column with buffer A, B**.** Plasmid DNA was eluted from the column with 5 mL elution buffer and precipitated with 0.7 volumes of isopropanol and collected by centrifugation at 20 000 x g for 18 minutes at 4 °C. The resulting plasmid DNA pellet was washed with 5 mL 75% EtOH and centrifugation at 20 000 x g for 18 minutes at 4 °C. The washed plasmid DNA pellet was then allowed to air-dry for 10 minutes and finally dissolved in 200–500 µL water or TE buffer.

2.5 RNA extraction

RNA extraction from transduced or genome edited MEL or human $CD34^+$ progenitor cells was performed using the Trizol kit (*Invitrogen, Thermo Fisher Scientific*) according to manufacturer's instructions. Briefly, ≥2 x 10^6 cells were lysed in 1 mL Trizol reagent and left at RTemp for 10 minutes. Samples were mixed with 200 µL chloroform, allowed to stand for 5 minutes at RTemp and centrifuged at 12 000 x g for 10 minutes at 4 °C. After centrifugation, three phases are formed; the top aqueous phase, an interphase and the bottom phenol-chloroform organic phase, which contain the RNA, DNA and proteins, respectively. The aqueous phase was carefully removed and transferred to a 1.5 mL RNase-free mini-tube. RNA was then precipitated by addition of 0.5 mL isopropanol, incubation for 10 minutes at RTemp and centrifugation at 12 000 x g for 10 minutes at 4 °C. RNA pellets were then washed twice with cold 75% EtOH and centrifugation at 12 000 x g for 10 minutes at 4 °C. RNA was allowed to air-dry for 10 minutes and dissolved in 50 µL of RNase-free water (*Sigma-Aldrich*). RNA pellets were completely dissolved by incubation of samples at 65 °C for 10 minutes.

2.6 General cloning procedure

All the new plasmid constructs produced in this project were made by following standard cloning procedures as described below. Positive colonies were confirmed by appropriate restriction enzyme digestion and sequencing, expanded and cryopreserved.

2.6.1 *DNA digestion by restriction endonucleases*

Restriction enzyme digestions were performed according to manufacturer's instructions (*NEB*). In general, 1–2 µg of DNA was used per digestion, 1/10 volume of appropriate 10 x NEB buffer, 1 µL of

the restriction enzyme of choice (2–10 units/µL) and made up to a 20 µL final volume with water. Bovine serum albumin (BSA) was added to the reaction when recommended by manufacturer. Restriction digest was performed at the optimal temperature of each enzyme for 30 minutes and if is necessary, followed by heat-inactivation of the enzyme, as recommended by manufacturer.

2.6.2 *Blunting ends using T4 DNA polymerase*

T4 DNA polymerase (*NEB*) was used for creating blunt-ends in DNA fragments with 5' or 3' overhangs, by exploiting the enzyme's 5'→ 3' DNA synthesis activity (5´ overhang fill-in to form blunt ends) and 3´→ 5´ exonuclease activity (3´ overhang removal to form blunt ends). Briefly, the desired purified DNA fragments were obtained by separation by agarose gel electrophoresis and extraction using silica based columns (see section 2.4.2). Briefly, 1/10 volume of 10 x NEB buffer 2.1 and 1 unit of T4 DNA polymerase per 1 µg of DNA was added in a final reaction volume of 20 µL containing 100 µM dNTPS. The reaction was incubated at 12 °C for 15 minutes, supplemented with 10 mM EDTA and heat-inactivated at 75 °C for 20 minutes.

2.6.3 *De-phosphorylation of vector backbone*

Dephosporylation of 5'-ends of DNA was performed using Antarctic phosphatase (*NEB*) according to manufacturer's instructions. Briefly, 1/10 volume of 10x Antarctic phosphatase buffer was added to 1–5 µg plasmid DNA cut with any restriction enzyme in any buffer and 5 units of Antarctic phosphatase. The reaction mixture was incubated for 15 minutes at 37 °C for 5' extensions and blunt ends, and 60 minutes for 3' extensions. Antarctic phosphatase was heat-inactivated for 5 minutes at 70 °C.

2.6.4 *Extraction of DNA fragments*

Digested DNA was separated by agarose gel electrophoresis for extraction of fragments of interest with the NucleoSpin® Gel and PCR Clean-Up Kit (*Macherey-Nagel GmbH*) according to manufacturer's instructions. Briefly, excised DNA-containing agarose gel bands were lysed in NTI buffer at 55 °C for 10 minutes, loaded onto the silica-based membrane column, washed and DNA fragments eluted with 30 µL of 5 mM Tris/HCl, pH 8.5 buffers.

2.6.5 *Production of inserts by oligo-annealing*

Small inserts with specific ends (blunt ends, specific 3' or 5' ends) were produced by oligo-annealing. Initially insert sequences were synthesised by Metabion (*Steinkirchen, Germany*) as two complementary oligo-nucleotide strands (forward and reverse) with the appropriate extension on

their 5' or 3' end, based on the desired type of overhang or without extension if blunt ends are needed (Table 23). Annealing of the oligonucleotides pairs resulted in formation of the double strand insert with the overhangs of our choice. The oligo-annealing consisted of a mixture containing 100μM of each forward and reverse oligonucleotide, 1x NEB buffer 3.1 and made up to a 50 μL final volume with water. Annealing was performed on a Veriti® thermal cycler (*Invitrogen™, Thermo Fisher Scientific*) programmed to start with a dissociation step at 95 °C for 4 minutes, followed by the annealing step, where the temperature was set to slowly fall by 15 °C in two hour.

2.6.6 *Ligations*

The molar concentration of DNA fragments was quantified by gel electrophoresis based on band intensities compared to DNA standards and fragment sizes, using ImageJ software for image analysis. In general, ligations were performed using Quick Ligase (*Roche, Basel, Switzerland*), and T4 Ligase *(NEB)* with two vector:insert ratios of 1:3 and 1:6, according to the manufacturer's instructions.

2.6.7 *Cloning oligonucleotides into pLKO.I vector*

Oligonucleotides were designed according to published guidelines of the RNAi consortium library (*http://www.broadinstitute.org/rnai/public/resources/protocols*) for expression of shRNAs under RNA pol3 promoters and inserts were produced by oligonucleotide annealing as described in section 2.6.6. Annealed oligonucleotides were specifically designed to create compatible ends for *Eco*RI (5') and *Age*I (3') restriction sites. The pLKO.I_shRNA expression vector was prepared to receive the desired oligonucleotides by digestion with *Eco*RI and *Age*I to remove the "Stuffer" fragment and purified by agarose gel electrophoresis / extraction as described in section 2.6.4. Annealed oligonucleotides (2 μL out of 50 μL annealing mixture) were ligated with 20 ng of pLKO.I vector and used to transform 50 μL NEB Turbo chemically competent *E coli*, as described in sections 2.6.6 and 2.17.3, respectively. Positive clones were picked following spreading of transformed bacteria on LB agar plates containing 100 μg/mL ampicillin and identity confirmed by appropriate restriction enzyme digestion and sequencing (Seq_U6_FW primer, Table 22), expanded and cryopreserved.

2.6.8 *TA TOPO cloning: Production of pCR2.1HBB cDNA^{Normal} and pCR2.1HBB cDNA^{Aberrant} constructs*

Constructs that carry cDNA regions from the correctly and abnormally spliced IVS1-110-β-globin mRNA were cloned using the *TA cloning kit* (*Invitrogen™, Thermo Fisher Scientific*) as recommended by the manufacturer. Briefly, the two inserts were produced in a typical PCR reaction (50 μL) using

AmpliTaq Gold polymerase, which adds deoxyadenosine (A) at the 3' end of the PCR products, a pair of primers (GCB1 FW and NewB RV) that encompasses the Exon 1 and Exon II junction sites and as template ~100 ng of cDNA from an IVS1-110 patient (product^{Normal} 309 bp and product^{Aberrant} 328 bp). The ligation reaction was performed by using 2 µL of fresh PCR product, 2 µL of pCR2.1 Vector (25 ng/µL), 1 µL of 10x ligation buffer, 1 µL T4 DNA ligase (4 Weiss units) and made up to 10µL with water. The ligation reaction was allowed to take place overnight (~12 hr) at 14 °C. Then, 2 µL of the ligation products were used for the transformation of OneShot competent bacteria, as described in section 2.17.3, from which positive clones were selected on LB agar plates supplemented with 50 µg/mL of kanamycin or 100 µg/mL ampicillin. Positive colonies were confirmed with appropriate restriction enzyme digestion and sequencing, expanded and cryopreserved.

2.6.9 *Mutagenesis to generate the MA821 HBB$^{IVSI-110\ (G>A)}$ LV*

Insertion of the HBB$^{IVSI-110\ (G>A)}$ mutation in the MA821 LV (GLOBE) construct was performed using the QuikChange Lightning Site-Directed Mutagenesis Kit according to the manufacturer's instructions (*Agilent Technologies*). Briefly, specific mutagenic oligonucleotides (IVS1-110_Mut_ FW and RV) were designed according to manufacturer's recommendations (25–45 bp with a melting temperature >=78 °C) harbouring the HBB$^{IVSI-110\ (G>A)}$ mutation (see Table 22). Primers were diluted to working concentrations (125 ng/µL). Mutagenesis reactions was prepared at 50 µL final volume, comprising 1x reaction buffer, 100 ng of each primer, 1 µL of dNTPs mix, 1.5 µL of QuikSolution reagent, 25 ng plasmid DNA (MA821) and topped-up with water, before the addition of 1 µL of QuikChange Lighting enzyme. Mutagenesis reaction cycle conditions start with denaturation at 95 °C for 2 minutes followed by 18 cycles of: denaturation at 95 °C for 20 seconds, primer annealing at 60°C for 10 seconds and extension at 68 °C for 5 minutes (30 sec/Kb of plasmid length), and a final extension step at 68 °C for 5 minutes. Following the cycling reactions, the product was treated with 2 µL *Dpn*I enzyme and incubated for 5 minutes at 37 °C. The DpnI endonuclease (target sequence: 5´-Gm6ATC-3´) is specific for methylated and hemi-methylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. Then, 6 µL of nicked mutated DNA vector was used for the transformation of XL10-Gold ultra-competent bacterial cells (see section 2.17.4).

2.7 DNA sequencing

Purified plasmids and PCR products were sequenced using BigDye Terminator v1.1 Cycler sequencing kit (*Applied Biosystems, MA, USA*) and the reaction performed on a Tgradient Thermocycler (*Biometra GmbH, Goettingen, Germany*) using the following reaction cycle conditions: 1 minutes at

96 °C followed by 25 cycles of, 10 seconds at 96 °C, 5 seconds at 50 °C, 4 minutes at 60 °C and finally a hold at 15 °C. Sequencing reaction master mix consisted of 4 µL BigDye v1.1, 4 µL 5 x BigDye buffer, 4 µL 5 x GC-RICH solution (*Roche, Basel, Switzerland*), 250 nM primer (Table 22), and 50 ng and 800 ng of purified PCR product and plasmid as template, respectively, in a final reaction volume of 20 µL. DNA sequencing products were purified using Performa® DTR Gel Filtration Cartridges Performa® DTR Gel Filtration Cartridges (*Edge Biosystems, Maryland, USA*), according to manufacturer's instructions. Sequencing results were analysed on a Hitachi 3031xl Genetic Analyzer with Sequence Detection Software version 5.2 (*Applied Biosystems, MA, USA*).

2.8 Construction of specific genome editing designer nucleases and dsEGFP reporters

2.8.1 *Creation of specific transcription activator like-effector nucleases (TALENs)*

Transcription activator like-effector nucleases (TALENs) were assembled using the Golden Gate assembly platform kit with the four modified Level III destination vectors to express functional nuclease monomers as described [227]. The target sequences were chosen to fulfil the two main TALEN pair designing rules: a) optimal spacer between the two TALEN monomers: 12–15 bp and b) thymidine (T) should precede the 18-nt target sequence. Based on the above requirements we designed two TALEN pairs that target the region adjacent to the $HBB^{IVSI\text{-}110(G>A)}$ mutation, two left monomers (HBB TALEN L1 and L2) and a common right monomer (HBB TALEN R1). In addition, three modified versions of the HBB TALEN R1 monomer, R2, R3 and R4, were prepared, in which specificity for the binding sequence was increased by substituting 2, 4 and 6 NN RVD modules, respectively, which are able to bind to guanine and adenine with NK RVD modules, which preferentially bind only guanine but with lower affinity [182]. A schematic representation of designer nucleases is shown in (Results Figure 47). The sequences of the TALEN RVDs are shown in Table 1. Generally, the Golden Gate custom assembly method of TALEN monomer constructs is completed in two cloning rounds. The assembly method is based on Type SII restriction enzyme sites, *Bsa*I and *Bpi*I, where the cleavage of DNA occurs on a different site than the DNA restriction site creating 4 bp overhangs. This allowed the development of a plasmid library kit, in which the TALE modules (RVD) are designed to create unique overlapping overhangs according to their position on the TALE repeat allowing the formation of pre-designed long arrays for TALENs [187,227]. The library consists of 80 Level I constructs, three Level II (A5, 5B and BC) and four Level III AC vectors, holding the last 0.5 RVD half repeat (HD or NG or NI or NK). Briefly, the appropriate Level I constructs, which encode the TALE repeat for specific RVD (Figure 16), were chosen from a library based on their position on the TALE repeat and were cloned in three Level II vectors (A5 (1–5 repeats), 5B (6–10 repeats) and BC (11–17 repeats),

each in three independent *Bsa*I cut/paste-ligation reactions. *Bsa*I cut / paste - ligation reaction includes: digestion at 37 °C for 5 minutes, ligation at 20 °C for 5 minutes), followed by 10 minutes at 50 °C and a 10 minute 80 °C heat inactivation step. The cut (*Bsa*I) / paste - ligation reaction comprises 1 µL Level II vectors (150 ng/µL), 1 µL from each Level I module (80 ng/µL), 1.5 µL *Bsa*I (NEB), 0.5 µL T4 ligase (NEB), 2 µL ligase buffer and top up to 20 µL with water. Constructs were transformed into DH5α *E coli* competent cells, as described in section 2.17.3 and plasmid DNA was extracted from positive (spectinomycin resistant) colonies. All Level II constructs confirmed by restriction test digestion with *Bpi*I expected bands: A5:~500bp 5B: ~650bp BC: ~750bp, were cloned together to the final Level III destination vector (AC) with a second round of cut/paste-ligation reactions as describe above, using *Bpi*I as restriction digest enzyme, 3 µL of each Level II constructs (A5, 5B and BC, 150 ng/µL) and 1 µL Level III AC final vector (150 ng/µL). The four versions of the final AC Level III vectors includes the final half-TALE repeat (17.5th RVD: HD or NG or NI or NK), CMV promoter, HA-taq, nuclear localisation signal (NLS), N' and C' terminal of TALE repeat, the wild type *Fok*I endonuclease cleavage domain, the poly A signal and kanamycin resistance (*Kan*^{R}) gene (Figure 16). Final ligation products were transformed into DH5α cells and spread on Agar LB plates with kanamycin antibiotic selection. Positive clones were picked and confirmed by restriction digest (*Pvu*II / *Hinc*II: ~2000bp, ~1700bp, ~1200bp, ~950bp, 350bp and *Bpi*I: uncut) and Sanger sequencing using TAL FW and RV primers (Table 22). Sequencing results were scrutinised using the ExPasY online tool (http://web.expasy.org/translate/), from which the correct order of the RVD of each TALEN construct was confirmed. Expression and integrity of the TALENs were confirmed by immunoblots of protein extracts from single transfected HEK293T cells with each TALEN monomer construct (N-terminal HA tag detection) (Figure 50).

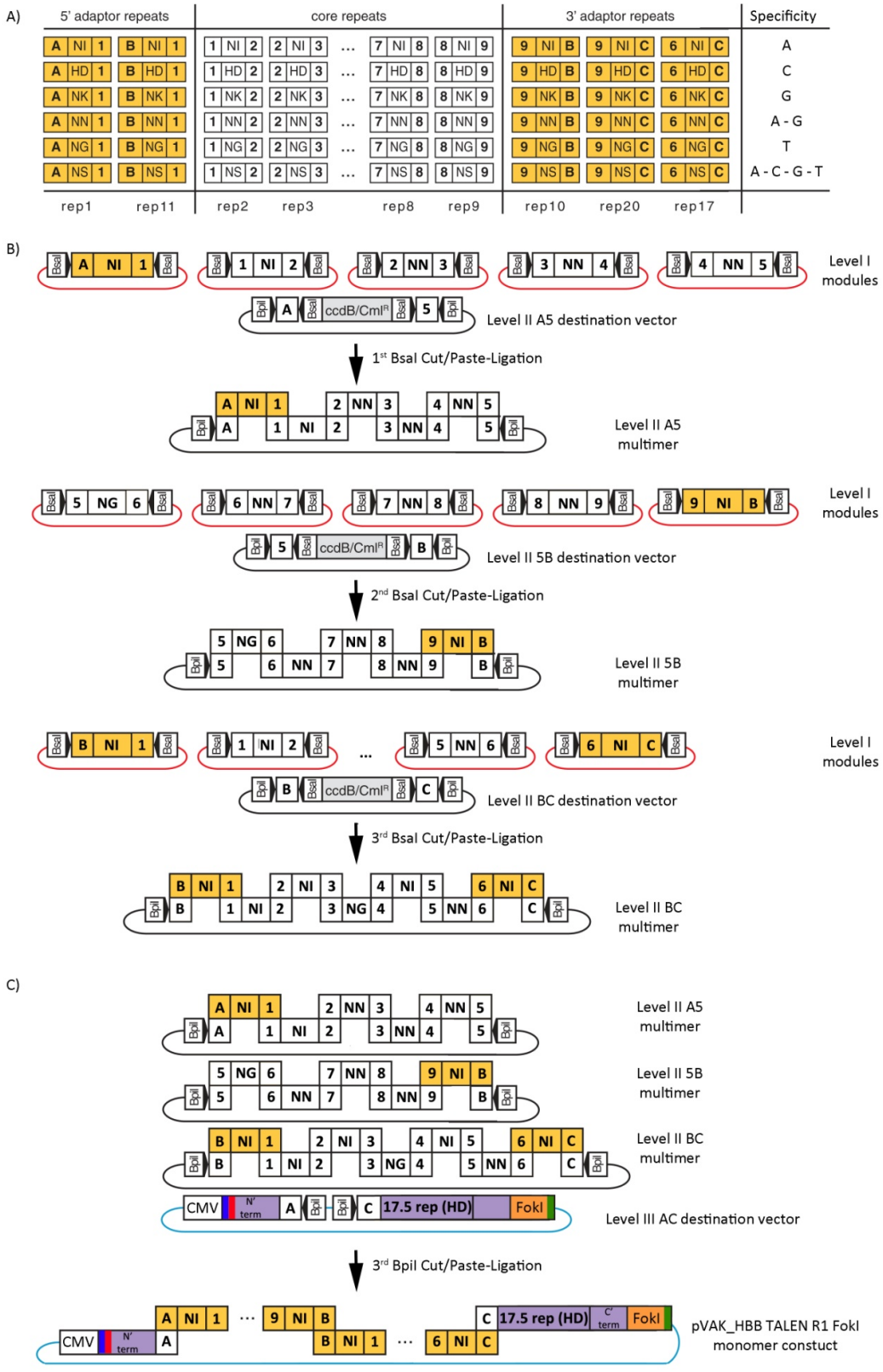

A)
5' adaptor repeats
core repeats
3' adaptor repeats
Specificity
A
C
G
A - G
T
A - C - G - T
rep1
rep11
rep2
rep3
rep8
rep9
rep10
rep20
rep17
B)
Level I modules
Level II A5 destination vector
1st BsaI Cut/Paste-Ligation
Level II A5 multimer
Level I modules
Level II 5B destination vector
2nd BsaI Cut/Paste-Ligation
Level II 5B multimer
Level I modules
Level II BC destination vector
3rd BsaI Cut/Paste-Ligation
Level II BC multimer
C)
Level II A5 multimer
Level II 5B multimer
Level II BC multimer
CMV
N' term
17.5 rep (HD)
FokI
Level III AC destination vector
3rd BpiI Cut/Paste-Ligation
C' term
pVAK_HBB TALEN R1 FokI monomer constuct

Figure 16: GoldenGate TALEN assembly platform.
Schematic representation of HBB TALEN R1 monomer (17.5 repeat) assembly using the GoldenGate platform. A) Level I module library categorized as 5' and 3' adaptors (yellow boxes) and core TALE repeats (white boxes) and their position on the TALE repeat. Distinct overlapping overhangs of repeat modules are designated in bold font (A, B, C. 1 – 9). Specificity of each RVD is designated on the right. B) Assembly of the three intermediated Level II TALE multimers in a pre-defined order, A5 (1st -5th repeat), 5B (6th – 10th) and BC (11th – 17th repeat) with 1st, 2nd and 3rd *Bsa*I cut/paste-ligations, respectively. Level I TALE modules are flanked with two *Bsa*I restriction sites, whereas Level II destination vectors have two *Bsa*I and *Bpi*I restriction sites, designated in white boxed with black arrowheads pointing to their cleavage sites. The grey box in the Level II vectors represents a Gateway cassette with ccdB gene, for the positive selection of recombinant clones and chloramphenicol (CmlR) antibiotic resistance gene. C) Assembly of the final pVAX_HBB TALEN R1 *Fok*I construct, in which the three intermediated Level II TALE multimers (A5, 5B and BC) were cloned in order to the final Level III vector AC with 1x *Bpi*I cut/paste-ligation reaction. Level III destination vectors have two *Bpi*I restriction sites, designated in white boxed with black arrowheads pointing to their cleavage sites. The final AC Level III vectors which includes the final 17.5th -TALE repeat (HD), CMV promoter, HA-taq (blue), nuclear localisation signal (NLS) (red), N' and C' terminal of TALE repeat (purple), the wild type *Fok*I endonuclease cleavage domain (orange), and the poly A signal (green). The lines connecting the boxes represents the backbone of each construct with colour indicating the antibiotic resistance of each construct (red: ampicillin, black: spectinomycin, blue: kanamycin). Modified image from [187].

2.8.2 *Creation of specific guide RNA for the CRISPR-Cas9 system*

Target sequence specific gRNAs were chosen based on the position of the PAM sequence (5'-NGG-3') relative to the mutation. The PAM sequence should be at the 3' end of the gRNA and the DSB will take place ~3–4 bp from this position. In the case of the $HBB^{IVSI-110}$ only one gRNA fulfilled the above requirements. The gRNA inserts were produced by oligonucleotide annealing, specifically designed to create compatible overhangs for *BsmB*I restriction sites (Table 23). The pLML3636 gRNA vector was prepared by restriction digestion with *BsmB*I and gRNA annealed products were ligated into the vector. DH5α competent bacteria were transformed with ligation products, as described in section 2.17.3 and grown on LB agar plates with ampicillin selection. Positive clones were confirmed by sequencing with CMV FW primer (Table 22). A schematic representation of the $HBB^{IVSI-110}$ RGEN target sequence is shown in Figure 47 in result section. Sequences of gRNAs are shown in Table 28.

2.8.3 *Production of $HBB^{IVSI-110}$ dsEGFP reporter construct*

The $HBB^{IVSI-110}$ dsEGFP reporter construct was used for the assessment of targeted disruption efficiencies of $HBB^{IVSI-110}$ specific designer nucleases using a plasmid-based assay via flow cytometry. The target binding sequence was inserted between the ATG and the 5'-end of a destabilized enhanced green fluorescent protein (dsEGFP) gene into the plasmid pLV.CMV.dsEGFP [228]. The $HBB^{IVSI-110}$ insert was prepared by oligonucleotide annealing, which were specifically designed to create compatible 5' overhangs for *Pac*I and *Age*I restriction sites (Table 23). pLV.CMV.dsEGFP reporter plasmid was prepared by digestion with *Pac*I and *Age*I restriction enzymes and ligated to annealed oligonucleotides containing the $HBB^{IVSI-110}$ mutation. DH5α competent *E. coli* cells were then transformed with ligation products as described in section 2.17.3 and grown on LB agar plates with Ampicillin selection. Positive clones were confirmed by Sanger sequencing with #1009 CMV Fw primer (Table 22).

2.9 Polymerase chain reaction (PCR)

In general, 100–200 ng of gDNA (or cDNA) were used per PCR reaction, 500 nM of each primer FW and RV (Table 22), 200 nM dNTPs, the 1 x of the appropriate buffer and made up to a final volume of either 25 µL or 50 µL with water. $MgCl_2$ was also added when needed to the reaction according to manufacturer's instructions.

2.9.1 *AmpliTaq GOLD PCR reaction*

AmpliTaq Gold PCR reaction was used for the production of PCR products with deoxy-adenosine (A) at the 3' end of the PCR products making them suitable for cloning into the TA plasmid vector as

described in section 2.6.8. AmpliTaq Gold PCR reaction consisted of 200ng gDNA/cDNA as template, 200 µM dNTPs, 0.5 µM of each primer, 1x PCR Gold Buffer , 1–4 mM $MgCl_2$, 1.25 Unit of AmpliTaq Gold Polymerase and made up to a final volume of 50 µL. Default cycling conditions were used. Note this polymerase requires an activation step of 10 minutes at 95 °C.

2.9.2 *Phusion Polymerase PCR reaction*

Phusion PCR reactions were used for the production of blunt-ended PCR products for their cloning to the TOPO 4 Blunt-ended vector (section 2.23.1), and destined for Sanger sequencing or the assessment of targeted disruption efficiency by T7E1 assay. The Phusion PCR reaction consisted of 200ng gDNA as template, 200 µM dNTPs, 0.5 µM of each primer, 1x GC Phusion Buffer ,1 Unit of Phusion Polymerase in a final volume of 50 µL. Default cycling conditions were used using specific annealing temperatures where designated (Table 1).

2.10 Purification of PCR products

2.10.1 *Silica-based PCR purification Qiagen Kit (Sequencing and T7E1 assay)*

PCR products destined for Sanger sequencing or T7E1 assay were purified using the Qiagen PCR purification kit (*Qiagen, Hilden, Germany*) according to manufacturer's instructions. Briefly, PCR products were made up to 100 µL with water, then 5 x volumes of PB binding buffer was added, and the resultant mixture loaded onto the QIAquick column. Columns were centrifuged for 1 minute at 17 900 x g. Columns were then washed with 750 µL of PE washing buffer, centrifuged for 1 minute at 17 900 x g followed by a second centrifugation to remove residual PE buffer. Finally, 25–50 µL of HPLC grade water was added to the centre of the QIAquick membrane, incubated for 1 minute at RTemp and the PCR products eluted into a new 1.5 mL microfuge tube with 1 minute centrifugation at 17 900 x g.

2.10.2 *ExoSap-IT PCR clean-up Kit*

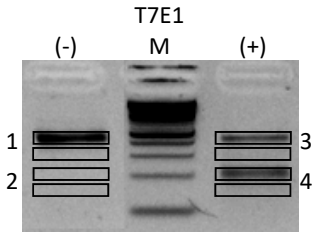
The ExoSap Kit (*Affymetrix, Thermo Fisher Scientific*) was used for the purification of PCR products produced from MEL MA821 $HBB^{IVSI-110}$ cell clones (section 2.23.2), meant for Sanger sequencing. In general, ExoSap-IT is a single step exonuclease I and Shrimp Alkaline Phosphatase (SAP) treatment of PCR products that eliminates unincorporated primers and dNTPs, which if not treated will impair the Sanger sequencing reaction. ExoSap treatment was done according to manufacturer's instructions. Briefly, 4 µL of Phusion PCR products were mixed with 1.5 µL of ExoSap-IT in a final volume of 7 µL made up with water. Samples were mixed and incubated for 30 minutes at 37 °C and heat-

inactivated at 80 °C for 20 minutes. ExoSap treated samples were ready to be used as templates for Sanger sequencing.

2.11 Assessment of designer nucleases' targeted disruption efficiency

2.11.1 *Assessment of targeted disruption efficiency on the genomic DNA level (T7E1 Assay)*

Specific primer pairs were designed to amplify the genomic region encompassing the target sequence (*HBB*, *HBD* and *CCR5*) of the designer nucleases using the optimized annealing temperature for a clear single PCR product (Table 1). Note: the *HBD* locus was used to evaluate possible off target effects of the *HBB* specific designer nucleases. In general PCR products of the locus of interest were denatured and re-annealed. In re-annealed PCR products from the genome edited population, normal and edited products form dsDNA with single-strand loops (heteroduplexes) due to low homology with the INDELs at the site of DSB. These loops constitute a substrate for cleavage by T7 endonuclease 1 (T7E1) enzyme (NEB) leading to the production of smaller sized products. Samples treated T7E1 (+) are resolved by agarose gel electrophoresis in parallel with the samples without T7E1 treatment (-) and based on the band densities quantified by ImageJ it can the percentage of cleaved heteroduplexes can be estimated, which is translated as targeted disruption efficiency of the designer nucleases after subtraction of background levels from nuclease-free negative control samples (pUC118). Briefly, 100 ng of gDNA was used as template for the Phusion PCR reaction. Amplicons were purified using the Qiagen PCR purification kit (*Qiagen*). Then, 300 ng of purified PCR products were heat-denatured for 5 minutes at 95 °C, re-annealed at 30 °C with a decreasing temperature rate of -0.05 °C /sec and held for 5 minutes at 4 °C. Re-annealed PCR products were divided into two tubes (150 ng each) and in one 6 units of T7 endonuclease I was added. Samples with and without T7E1 were incubated for 10 minutes at 37 °C. Then, samples were separated on 2.5% agarose gels at 120 V for 15 minutes. Intensities of uncleaved and cleaved bands were quantified by ImageJ software. Targeted disruption efficiency was calculated using the following formula (Figure 17).



Band 1 – background 1 = a Band 3 – background 3 = d

Band 1 – background 2 = b Band 4 – background 4 = e

% cleavage w/o T7E1 Cab= b/(a+b) % cleavage with T7E1 Cde= e/(d+e)

Proportion of targeted disruption = Cde – Cab

Percentage of targeted disruption: 100 x (Proportion of targeted disruption)

Figure 17: Quantification of TALEN and RGEN targeted disruption by T7E1 assay.
T7E1 assay with agarose gel electrophoresis was used for the assessment of targeted disruption efficiencies of specific designer nucleases (TALEN or RGEN). Heteroduplex of PCR products without (-) and with (+) T7E1 treatment were resolved on agarose gels and bands densities quantified with Image J software. This was achieved by measuring the raw integrated intensities of a defined region. Percentage of cleavage in each lane was calculated after subtracting first the background density of each band. Finally, the percentage of targeted disruption was calculated by subtracting the percentage of cleavage in the sample without T7E1 (-) (Cab) from the percentage cleavage with T7E1 (+) (Cde).

Table 1: Primer pair sets for Phusion PCR amplification of designer nucleases target region prior to T7E1 Assay.

Primer	Target Locus	Annealing temperature (°C)	Product (bp)
GCB1 FW / HBB EX2.2 RV	HBB $^{IVSI-110}$	68	432
GCB1 FW / 109 RV	HBB $^{IVSI-110}$	69	542
GCB1 FW / HBD EX2.1 RV	HBD of Homologous region of HBB $^{IVSI-110(G>A)}$ site	68	431
CCR5 FW /CCR5 RV	CCR5	60	292

2.11.2 *Assessment of targeted disruption efficiency on HBB$^{IVSI-110}$-GFP reporter construct (flow cytometry)*

Targeted disruption efficiency of the HBB$^{IVSI-110}$ specific designer nucleases was assessed by flow cytometry using the pLV.CMV.HBB$^{IVSI-110}$ dsEGFP reporter plasmid (section 2.8.3). The HBB$^{IVSI-110}$ GFP reporter construct was designed to hold the HBB$^{IVSI-110}$ target sequence in-frame upstream from the dsEGFP cDNA sequence. Disruption of the HBB$^{IVSI-110}$ target sequence by specific TALENs or RGEN leads to a frameshift in the eGFP coding sequence. Loss of eGFP fluorescence, which is easily detected by flow cytometry, is inversely proportional to cleavage activity of the designer nuclease. At 48 hours post-transfection cells were harvested for flow cytometry analysis. Between 10^5 and 10^6 transfected cells were harvested (following trypsinisation) and centrifuged at 250 x g for 10 minutes. Each cell pellet was re-suspended in 2 mL cold PBS (2% FBS). The cell suspension was then divided into two flow cytometry tubes (1 mL/tube) and 1 µL of SYTOX Red dead cell stain (*Life Technologies,*

Thermo Fisher Scientific) was added before all samples were incubated for 15 minutes on ice in the dark prior to flow cytometry analysis.

2.12 RNA Analysis

2.12.1 *DNase I treatment*

RNA extracted using TriZol™ kit (*Invitrogen™, Thermo Fisher Scientific*) was quantified spectrophotometrically and 1 µg was treated with 0.5 units of DNase I (*Invitrogen™, Thermo Fisher Scientific*) according to manufacturer's instructions. Briefly, 1 µg of RNA was mixed and incubated for 30 minutes at 37 °C with 0.5 units DNase I, 1 x DNase Buffer, made up to 10 µL final reaction volume with water. DNase I was then inactivated by the addition of 1 µL 25mM EDTA and incubation at 65 °C for 10 minutes.

2.12.2 *Quantification of mRNA expression by RT-qPCR*

Gene expression quantification was performed in a two-step reverse-transcription qPCR. First, single-strand reverse transcription was achieved using the TaqMan Reverse transcription PCR kit (*Applied Biosystems, MA, USA*) following the manufacturer's instructions.

2.12.2.1 *Reverse transcription PCR (1st step)*

A 250 ng aliquot of DNase I treated RNA was reverse transcribed in a final reaction volume of 10 µL consisting of 2.5 µM random hexamers, 0.4 U/µL RNase inhibitor, 500 µM deoxyribonucleotides (dNTPs), 5.5 mM $MgCl_2$, 1 x TaqMan buffer and 1.25 U/µL MultiScribe™ Reverse Transcriptase. RT-PCR cycle conditions began with 10 minutes at 25 °C (primer binding), 30 minutes at 48 °C (reverse transcription) and 5 minutes at 95 °C (enzyme inactivation). The resulting cDNA samples were diluted 4 fold with RNase-free water (*Sigma-Aldrich*) to a final concentration ~6.25 ng/µL and 2 µL were used for each qPCR reaction.

2.12.2.2 *Quantitative PCR (2nd step)*

Gene expression quantification was performed using the 7900HT Fast Real Time PCR System. Each sample was run in duplicate employing the SYBR Green detection system (*Applied Biosystems*) and triplicate in probe-based multiplex qPCR reactions (*Qiagen*). Non-template negative controls (NTCs) were always included.

2.12.2.2.1 SYBR Green method

Relative quantification of expression of human *HBB* and murine *Hba1/2* was achieved using SYBR Green I chemistry containing a ROX passive reference dye (*Applied Biosystems*) and $2^{-\Delta(\Delta Ct)}$ analysis method using murine *Actb* expression as a housekeeping reference gene relative to the non-treated controls. Similarly, *HBB* expression can be measured relative to the degree of cell erythroid differentiation, using *Hba1/2* expression as a reference. This approach was specifically used for the normalization of absolute quantities of the two *HBB* variants, normal and aberrant, (measured by multiplex RT-qPCR; see below) upon differentiation - normalised total β-globin expression in the differentiated edited MEL MA821 clones relative to the non-edited cells.

Briefly, 12.5 µL of 2 x SYBR Green PCR Master Mix , 300 nM of each forward and reverse primers (*Metabion International AG*, *Planegg, Germany*), and 2 µL of (~6.25 ng/µL) 4x diluted template cDNA were combined per reaction in a total volume of 25 µL. Default cycle conditions were used for *Actb* expression, starting with activation of the polymerase at 95 °C for 10 minutes, followed by 40 cycles of two steps, denaturation and annealing/extension, at 95 °C for 15 seconds and 60 °C for 1 minute, respectively. The annealing/extension step for HBB and Hba reactions was set to 65 °C for 1 minute to avoid non-specific detection of the murine Hbb transcripts or the formation of primer-dimer artefacts, respectively. Contamination and formation of primer-dimers was assessed by performing dissociation curve analysis at the last stage of the qPCR reaction, in which samples were heated at 95 °C for 15 seconds, cooled down to 60 °C for 15 seconds and finally heated again to 95 °C at a ramp rate of 2% maximum speed for 15 seconds with continuous fluorescence detection. Gene-specific reverse transcription primer pairs are shown in Table 2. Relative quantification via the $2^{-\Delta(\Delta Ct)}$ method was measured using the following equations:

$$\Delta Ct = Ct_{Target} - Ct_{Reference}$$

$$\Delta\Delta Ct = \Delta Ct_{Sample} - \Delta Ct_{Control}$$

Fold change of expression level relative to control: $2^{-\Delta(\Delta Ct)}$

Where target and reference corresponds to the Ct values of the gene of interest and the housekeeping gene (murine Actb) or differentiation gene (murine Hba1/2), difference in expression of the gene of interest is calculated as (ΔΔCt) relative to the non-treated control, with measures of $2^{-\Delta(\Delta Ct)}$ in the final outcome expressed as fold-change of the expression of samples relative to the un-treated control.

2.12.2.2.2 Probe-based Multiplex method

Relative quantities of aberrantly and correctly spliced HBB mRNA were measured in a multiplex RT-qPCR using a plasmid-based calibration standard curve employing known quantities of pCR2.1_HBB_cDNA^{normal}_cDNA^{aberrant} ($1x10^{-1}$, $1x10^{-2}$, $1x10^{-3}$, $1x10^{-4}$ and $1x10^{-5}$ ng which corresponds to $2.01x10^{5}$, $2.01x10^{4}$, $2.01x10^{3}$, $2.01x10^{2}$ and $2.01x10^{1}$ molecules, respectively). Briefly, multiplex PCR reactions is consisted with 6.25 µL of 2x Qiagen multiplex master mix (Qiagen), 900 nM of each primer (HBB EX1_FW_3 and HBB EX2_RV_1), 250 nM of each specific probe (IVSI-110_MGB_VIC and wtHBB_ZNA_FAM) and as template 2 µL of plasmid DNA SC dilutions or (~6.25 ng/µL) 4x diluted sample cDNA and topped - up with water to 12.5 µL final volume. The real-time PCR reaction started with activation of the polymerase at 95 °C for 15 minutes, followed by 40 cycles of denaturation at 95 °C for 30 seconds and annealing / extension at 60 °C for 1 minute, respectively.

Based on the plasmid standard curve, absolute quantities of each variant mRNA were calculated leading to the assessment of the percentage of each variant in the total population as also the percentage difference of each variant relative to non-treated –ve controls. Specifically for the correlation of particular INDELs with splicing restoration at the clonal level of MEL MA821 HBB$^{IVSI-110}$ VCN1, we normalised the absolute quantities of the two *HBB* variants measured by multiplex RT-qPCR, on relative quantities of total *HBB* expression measured comparative to the degree of cell erythroid differentiation using the $2^{(-\Delta\Delta Ct)}$ method described above (section 2.12.2.2.1) in which murine Hba1/2 mRNA is used as a reference. In this way, absolute quantities of each variant are normalised on cell erythroid differentiation, providing a more accurate assessment of the actual level of splice restoration in each edited clone compared to the non-edited. For the quantification of the two *HBB* variants the following formulas were used:

Percentage of each HBB variant in the total HBB mRNA population

a) % HBB^{Normal} mRNA = 100 x (abs. quant. HBB^{Normal} mRNA/ abs. quant. HBB^{Normal} + HBB^{Aberrant} mRNA)
b) % HBB^{Aberrant} mRNA = 100 - % HBB^{Normal} mRNA

Percentage of change each HBB variant relative to non-treated negative controls

a) HBB^{Normal} mRNA % change = 100 x (% HBB^{Normal} mRNA $_{treated}$/% HBB^{Normal} mRNA $_{non\text{-}treated}$)
b) HBB^{Aberrant} mRNA % change = 100 x (% HBB^{Aberrant} mRNA $_{treated}$/HBB^{Aberrant} mRNA $_{non\text{-}treated}$)

Normalised quantities of each HBB variant on differentiation-normalised total *HBB* expression

a) Fold change expression of differentiation-normalised total HBB expression:
   $\Delta Ct = Ct_{total\ HBB} - Ct_{Hba}$
   $\Delta\Delta Ct = \Delta Ct_{edited\ clone} - \Delta Ct_{non\text{-}edited\ clone}$
   Fold changes of differentiation normalised HBB expression relative to non-edited control = $2^{(-\Delta\Delta Ct)}$
b) Norm. Quant. HBB^{Normal} = (abs. quant. HBB^{Normal} mRNA) x ($2^{(-\Delta\Delta Ct)}$)
c) Norm. Quant. HBB^{Aberrant} = (abs. quant. HBB^{Aberrant} mRNA) x ($2^{(-\Delta\Delta Ct)}$)

Table 2: Primer and probe combination and annealing temperatures for RT-qPCR.

Method	Primer	Transcript	Annealing temp
SybrGreen	hHBB_EX1_FW_1 / hHBB_EX1_RV_2_A	total human HBB on Exon 1	65 °C
	hHBB_FW_EX2_B hHBB_EX2.3_RV_B	total human HBB on Exon 2	65 °C
	mouse β-Actin Fw / Rv	murine Actb	60 °C
	mHba-a1/a2 FW / RV	murine Hba1/2	65 °C
Qiagen Multiplex RT-qPCR	hHBB_EX1_FW_3 / hHBB_EX2_RV_1	normal and aberrantly spliced HBB variants	60 °C
	wtHBB_Probe_ZNA		
	IVS1-110_MGB_Probe		

2.13 Construction of pCR2.1_HBB_cDNA$^{Normal/Aberrant}$ standard curve

Two independent plasmid constructs containing the cDNA fragments encompassing the region of the exon 1 / exon 2 junction site of a correctly and aberrantly spliced mRNA were prepared by TA-cloning (section 2.6.8). In order to increase accuracy of the standard curve, cDNA^{Aberrant} fragment was cloned from pCR2.1_HBB_cDNA^{Aberrant} to the pCR2.1_HBB_cDNA^{Normal} construct in the same orientation. First, *EcoR*V and *Xho*I restriction sites were incorporated into the cDNA^{Aberrant} fragment via Phusion PCR of the pCR2.1 HBB cDNA^{Aberrant} construct using IVSI-110_Cloning FW and RV primers (Table 22). Then, restriction digests of both HBB_ cDNA^{Aberrant} PCR product and pCR2.1_HBB_cDNA^{Normal} construct with *EcoR*V and *Xho*I followed by standard cloning procedures (ligation / transformation; section 2.6.6) gave the pCR2.1_HBB_cDNA^{Normal}_cDNA^{Aberrant} construct, which was confirmed by Sanger sequencing using M13_Fw primer (Table 22). A series of 10-fold serial dilutions of pCR2.1_HBB_cDNA^{Normal}_cDNA^{Aberrant} of known concentration (0.1, 0.01, 0.001, 0.0001, 00001 and 0.000001 ng/μL corresponding to $2.01x10^{6}$, $2.01x10^{5}$, $2.01x10^{4}$, $2.01x10^{3}$, $2.01x10^{2}$ and $2.01x10^{1}$ molecules/μL, respectively) were made and from which 2 μL were loaded per reaction. For the preparation of the above standard curve plasmid mass was quantified using the following formula:

$$m = s \times (1.096^{-21})$$

where (m) is the mass of the plasmid (g/ μL), (s) the size (bp). The appropriate final plasmid concentrations where calculated using the following formula:

$$Ci \times Vi = Cf \times Vf$$

where (C) and (V) represent concentration (ng/μL) and volume (μL), respectively, and (i) and (f) stand for initial and final, accordingly. Each SC dilution was assayed in triplicate. In addition, external equations of each variant were made based on the observed Ct values and known number of molecules of the plasmid-based calibration standard curve (Figure 22). External equations were used

for the re-quantification of the data from samples that were run with a two construct SC in 1:1 ratio (pCR2.1 HBB cDNA^{Normal} and pCR2.1 HBB cDNA^{Aberrant}), which subsequently was shown to be not accurate (Figure 21). Importantly, in order for the comparison of the mRNA quantities measured with the external SC equations in treated samples and non-treated negative controls to be valid, all the samples had been run on the same reaction plate, to avoid errors due to difference in PCR efficiency between different reactions.

Standard curve external equations:

a) SC HBB^{Normal}: $y_{normal} = 4x10^{8}e^{-0.491x}$

b) SC HBB^{Aberrant}: $y_{aberrant} = 1x10^{9}e^{-0.526x}$

where (y) represents the quantities of the corresponding mRNA variants and (x) the observed Ct values of each samples.

2.14 Protein Analysis

2.14.1 *Preparation of protein samples for SDS PAGE electrophoresis*

Cells were lysed for 20 minutes on ice in radioimmunoprecipitation (RIPA) lysis buffer supplemented with protease inhibitor (*Roche, Basel, Switzerland*) at a cell concentration 1 x 10^6 cells/10 µL. Cell lysates were centrifuged at maximum speed for 10 minutes at 4 °C and the supernatant transferred to new 1.5 mL mini-tubes. Equal volume of 2 x sample buffer was added to the proteinaceous supernatants. Before loading samples on polyacrylamide gels, samples were denatured for 5 minutes at 100 °C. Protein extracts from 0.5–1 x 10^6 cells were resolved in parallel to molecular weight protein markers (*Nippon Genetics, Duren, Germany*) for protein size determination on SDS PAGE gels comprised of a top stacking gel and bottom resolving gel. The electrophoresis apparatus was assembled and gels were submerged in 1 x running buffer. Samples were resolved at constant 150 Volts for 1 hour. Details of PAGE gel concentrations used is shown in Table 3.

Table 3: Stacking and resolving gel solutions

Reagents	Resolving gel		Stacking gel
	12%	**15%**	
dH2O	3.44 mL	2.84 mL	3 mL
40% acrylamide	2.4 mL	3 mL	0.625 mL
1.5 M Tris pH 8.8	2 mL	2 mL	-
0.5 M Tris pH 6.8	-		1.260 mL
20% SDS	40 µL	40 µL	25 µL
30% APS	26.8 µL	26.8 µL	16.6 µL
TEMED	8 µL	8 µL	5 µL

2.14.2 *Blotting*

Proteins were transferred from the SDS PAGE gels onto nitrocellulose Parablot NCP membranes (*Macherey-Nagel GmbH*) using wet electrophoretic transfer. First, the transfer cassette was prepared in following order: sponge, 2 filter papers (*Macherey-Nagel GmbH, Duren, Germany*), protein-containing gel, nitrocellulose membrane, two filter membranes and sponge. All membranes and filter papers were pre-wetted in 1 x blotting buffer. The cassette was submerged in the BioRad apparatus tank filled with 1 x blotting buffer. Complete transfer of proteins to nitrocellulose membranes was achieved with electrophoresis at constant 300 mA for 2 hours with the device incubated on ice to prevent temperature fluctuations. Membranes were temporary stained with Ponceau Red solution (*Sigma-Aldrich, UK*), to confirm the successful transfer of the proteins and de-stained by repeated washes in dH2O or TBS tween.

2.14.3 *Immunoblotting and detection*

Membranes were blocked for 1 hour in blocking buffer (1% *BSA* (*Sigma-Aldrich, UK*) in TBS tween) on a rocking platform. Then, membranes were incubated with the primary antibodies at 4 °C on a roller. Primary antibody working concentrations are shown in Table 4. Membranes were washed four times in TBS tween for 10 minutes and incubated with the appropriate secondary antibodies (Table 4) for 1 hour on a rocking platform followed by another four 10-minute washes. Finally, membranes were covered with chemiluminescence staining buffer (*Lumisensor, GenScript, New Jersey, USA*) for 3 minutes in the dark. Protein bands were visualized / detected by the Bio-Rad Imaging system and band density quantification was measured with ImageLab Software 5.1 (*Bio-Rad Laboratories, Inc., CA, USA*).

The correction level of the $HBB^{IVSI\text{-}110\ (G>A)}$ mutation in treated samples (genome edited or transduced) was measured as the fold-change of the ratio of HBB/ Hba chain band intensities relative to the non-treated negative control or as the percentage relative to the MA821 HBB^{Normal} positive control. Stage of differentiation was measured as fold-change in the level of murine Hba chains. Murine Actb was used as a protein loading control.

Table 4: Antibodies.
Primary and secondary antibodies used for immunoblots with the appropriate working concentrations. Primary antibodies are diluted in 1% BSA / TBS-tween, with exception of mouse anti-Actb (AC-15) in which is it was diluted in 1.5 % milk /TBS-tween. Secondary antibody working concentrations were prepared in 3% milk / TBS – tween.

Primary antibodies	catalog number	Company	Working Dilution
Mouse monoclonal anti - human HBB (37-8)	sc-21757	Santa Cruz biotechnologies	1 : 1000
Rabbit polyclonal anti - murine Hba (H80)	sc-21005	Santa Cruz biotechnologies	1 : 1000
mouse monoclonal anti - murine-Actb antibody (AC-15)	A1978	Sigma Aldrich	1 : 10000
mouse monoclonal HA-probe Antibody (F-7)	sc-7392	Santa Cruz biotechnologies	1 : 10000
Secondary antibodies	**catalog number**	**Company**	**Working Dilution**
peroxidase-conjugated AffiniPure goat anti-mouse IgG (H=L)	115-035-003	Jackson ImmunoResearch Laboratories, PA, USA	1 : 10000
peroxidase-conjugated AffiniPure goat anti-rabbit IgG (H=L)	111-035-003	Jackson ImmunoResearch Laboratories, PA, USA	1 : 10000

Table 5: Immunoblot: Buffer and solutions

Buffer	Composition
1.5 M Tris pH 8.8	18.15 g Tris base 100 mL dH2O pH 8.8 with 6 N HCl
0.5 M Tris pH 6.8	6 g Tris base 100 mL dH2O pH 6.8 with 6 N HCl
RIPA buffer	150 mM NaCl 20 mM Tris-HCl, pH 7.4 1% NP40 0.1% SDS 0.5% sodium deoxycholate 5 mM EDTA
2X Sample Buffer	5 mL 0.25 M Tris pH 6.8 2 mL 10 % SDS 2.1 mL Glycerol 0.4 mL β-mercaptoethanol 0.5 mL 0.1 % bromophenol blue
10X Running Buffer	30.2 g/L Tris 144.2 g/L Glycine 1% SDS
10X Blotting Buffer	30.2 g/L Tris 144.2 g/L Glycine
1X Blotting Buffer	100 mL 10X Blotting Buffer 150 mL methanol 750 mL dH2O
10X TBS	87.7 g/L NaCl 12.1 g/L Tris, pH 7.4
1X TBS — 0.05% Tween 20 (TBST)	100 mL 10X TBS 5 mL 10 % Tween 20 895 mL dH2O

2.15 Separation of globin chains by reversed-phase liquid chromatography

2.15.1 *Instrument*

Globin chain separation was performed using a Shimadzu LC-20AD chromatographic system (*Shimadzu, Kyoto, Japan*), equipped with a CBM-20Alite system controller, a LC-20AD dual-plunger solvent delivery system, a SPD-M20A UV-Vis variable wavelength (190–800 nm range) detector, a SIL-20AC auto-injector, a CTO-20A column oven, and an LC-Solution work station interfaced with the Ver. 1.1 software package.

2.15.2 *Sample preparation*

Induced human $CD34^+$ cells were collected on day 5–7 post-differentiation. At least 1 x 10^6 cells were needed for analysis. Cells were pellet at 300 x g for 5 minute, washed once with 500 µL PBS and repeat centrifugation. Cell pellets were then lysed in a minimum final volume of 50 µL and concentration of 1x10^6 cells/50 µL in HPLC water and two rounds of freezing / thawing at -80 °C and on ice. Protein extracts were collected as a supernatant after centrifugation at 16 000 x g (maximum speed) for 10 minutes at 4 °C to remove cellular debris. Supernatants were transferred in 250 µL HPLC micro - inserts or HPLC vials (*Grace Vydac, USA*) and a minimum volume of 25–30 µL of lysate/sample was applied onto the Jupiter C18 column (*Grace Vydac, USA*) for analysis with injection volume equivalent to 0.5 x 10^6 lysed erythroid cells.

Reversed-phase high-performance liquid chromatography (RP-HPLC) separates peptides on the basis of their hydrophobicity by applying an increasing linear gradient of organic solvent (Phase A (0.1% TFA 4 mM NaOH in H_2O) to Phase B (2:1 acetonitrile:methanol (v:v)) to the chromatographic column. Samples were maintained at 15 °C and flow cell at 22 °C. The HPLC programme starts with a constant flow rate of 1 mL/min and at a baseline mobile phase (phase A at 44% and phase B at 56%) for 2min. Then, a gradient of organic phase B is increased reaching to 62% on the 8^{th} minute (haeme elution), 68% on 14^{th} minute (Gradient rate: 1%/min) and finally returned back to baseline mobile phase on the 16^{th} min. According to this protocol, haeme and each globin chain is released from the column and quantified at a specific time (Table 6) and displayed as peaks. The surface of the peaks corresponds to the amounts of the specific peptide. The expression levels of HBB-like chains are calculated as the ratio of the peak area of HBB-like/HBA –chains. Under normal conditions a balance between HBB and HBA chains is maintained leading to a ~0.9 HBB/HBA ratio. Changes in the expression of HBB-like chains in treated samples are analysed as the percentage change of the corresponding ratios compared to untreated controls.

Table 6: Retention time of peptides in HPLC analysis

Protein	Retention time
Haeme	7.500 min
Hs preβ	9.000 min
Hs β	9.700 min
Hs δ	10.800 min
Hs α	12.500 min
Hs Gγ	13.350 min
Hs Aγ	14.270 min

Percentage change = 100 x ((ratio β/α globin chains $_{treated}$)/ (ratio β/α globin chains $_{un\text{-}treated}$))

Estimation of HBB chain expression in genome edited subpopulations

This calculation method was applied on data from immunoblots and HPLC analysis, for the estimation of HBB chain increase in genome-edited transgenic MEL MA821 HBB$^{IVSI\text{-}110}$ VCN 1.9 cell line and HBB$^{IVSI\text{-}110}$-patient-derived hCD34^{+} cells, respectively. For the estimation of HBB chain expression in genome-edited subpopulations, we normalised the difference of HBB/HBA or Hba fold increase in genome-edited cells relative to the non-edited cells, on the percentage of genome-edited cells in the total treated population. In samples where no increase of HBB:HBA or Hba ratio was detected and/or cells were not edited (0% of *HBB$^{IVSI\text{-}110}$* targeted disruption), values without normalisation were used for the estimation. Results are given as the fold increase in the edited fraction plus 1 which correspond to the background levels produced in non-edited cells.

Example:

Table 7: Example of the estimation of HBB chain expression in genome edited cells.
Percentage of genome edited cells/*HBB$^{IVSI\text{-}110}$* targeted disruption is measured by T7E1 assay described previously. Fold increase of HBB/HBA or Hba in edited cells is measured from the globin chain quantification by western immunoblots or HPLC analysis as mentioned previously. HBB/HBA increased due to edited cells is calculated as the difference of the HBB/HBA or Hba fold increase in genome edited population relative to the non-edited control (HBB TAL R1/L1: 1.071426 – 1 = 0.0714). Extra HBB produced in edited fraction is calculated as the difference in HBB/HBA or Hba fold increase normalised on % of genome edited cells (HBB TAL R1/L1: 0.0714/0.1165=0.613). Therefore the total fold increase compared to non-edited cells is the extra HBB produced in edited fraction plus the background levels (1) (HBB TAL R1/L1: 0.612+1=1.613 HBB/HBA or Hba ratio fold increase relative to non-edited cells). In samples where no-increase of HBB:HBA or Hba ratio was detected (e.g. HBB TAL R1/L2 and HBB$^{IVSI\text{-}110}$ RGEN) and/or cells were not edited (e.g. HBB$^{IVSI\text{-}110}$ RGEN), values without normalisation were used for the estimation.

Samples	% *HBB$^{IVSI\text{-}110}$* targeted disruption	Fold increase by immunoblot/ HPLC analysis	Calculated HBB/HBA increased due to edited cells	Extra HBB produced in edited fraction	Total fold difference in edited fraction population
HBB TAL R1/L1	11.65	1.071426	0.071426	0.613099	1.613099
HBB TAL R1/L2	9.1	0.876994	n/a	n/a	0.876994
HBBI$^{IVSI\text{-}110}$ RGEN	-0.51	0.738353	n/a	n/a	0.738353
Non-edited cells	0	1	n/a	n/a	1

2.16 Flow Cytometry analysis

Flow cytometry was performed using a CyFlow Cube 8 v.6 channel instrument (*Partec, Sysmex, Germany*), equipped with a dual-laser system (488 nm blue and 632 nm red) and 6 optical parameters (Forward Scatter, Side Scatter and four fluorescence channels (FL1–FL4) with the Windows™ based CyView Cube 1.3.2.4 software. Data analysis was performed with FCS Express 5 Flow Cytometry software (*De Novo Software, CA, USA*).

Transfection efficiency, cell death and eGFP reporter gene-based targeted disruption assays were measured by flow cytometry. Transfected cells (5 - 10 x 10^5) were washed once with 1 mL PBS and cells recovered by centrifugation at 300 x g for 5 minutes. Cell pellets were re-suspended in 500 µL cold PBS, divided into two flow cytometry tubes, which contained 250 µL PBS and 250 µL PBS supplemented with 1x SYTOX Red dead cell stain (*Life Technologies, Thermo Fisher Scientific)*. Samples were incubated on ice in the dark for 10 minutes. A minimum of 10 000 cellular events were recorded.

2.17 Bacterial cultures

2.17.1 *E.coli strains*

Commercially available DH5α™ and One Shot® TOP10 competent cells (*Invitrogen™, Thermo Fisher Scientific*) and Turbo competent cells (*NEB*), XL10-Gold ultra-competent cells (*Agilent Technologies*) cells were used.

2.17.2 *Preparation of bacterial growth medium*

Lysogeny broth was prepared using Lowry Broth Powder (*Sigma-Aldrich*) according to the manufacturer's instructions. In general, 25 g of Lowry Broth Powder (*Sigma-Aldrich*) was added per 1 L of Millipore-purified water mixed and sterilized by autoclaving.

Solid LB growth medium was prepared with the addition of 1.5 - 2 g of agar (*Sigma-Aldrich*) per 1L of LB growth medium and sterilized by autoclaving.

2.17.3 *Transformation of DH5α™, One Shot® TOP10 and Turbo competent cells (NEB®)*

Competent bacteria were thawed on ice for 10 minutes and plasmid DNA (20 ng) or ligation product added to not more than 1/10 of the bacterial volume. The bacteria/DNA mixture was incubated for 10 minutes on ice and heat-shock at 42 °C for 40 seconds followed by 5 minutes incubation on ice. Then, 900 µL of SOCS (antibiotic-free) medium was added to the transformed bacteria and cultured

for 60 minutes in a shacking incubator at 37 °C. Transformed bacteria were centrifuged at 5000 x g for 2 minutes, re-suspended in 100 µL of LB medium, spread on LB Agar plates with the appropriate antibiotic selection (spectinomycin, 50 µg/mL; kanamycin, 50 µg/mL; ampicillin, 100 µg/mL). Plates were incubated overnight at 37 °C (~12hours).

2.17.4 *Transformation of XL10-Gold ultra-competent bacteria cells*

XL10-Gold ultra-competent *E coli* bacterial cells (*Agilent Technologies, USA*) were thawed on ice. A 45 µL aliquot in pre-chilled 14 mL BD falcon tube was used for each transformation to which was added 2 µL of β-mercaptoethanol, mixed gently and incubated on ice for 2 minutes. Then, 6 µL of the *Dpn*I-treated DNA was added to the bacteria, mixed gently and incubated for 30 minutes on ice. Bacteria were then transformed by heat-shock at 42 °C for 30 seconds, incubated on ice for 2 minutes, and cultured for 1 hour at 37 °C in 0.5 mL of pre-heated (42°C) SOCS medium. Transformed bacteria were then spread on LB–agar plates with ampicillin selection and cultured overnight at 37 °C.

2.17.5 *Growth of bacteria clones and plasmid preparation*

Single colonies were picked and cultured in 5 mL LB with the appropriate antibiotic selection (spectinomycin, 50 µg/mL; kanamycin, 50ng/mL; ampicillin, 100 ng/mL) in shaking incubator at 37 °C. Plasmid DNA was extracted using a plasmid mini-prep protocol (section 2.4.1). All clones were confirmed initially by a restriction enzyme digest test (section 2.6.1) followed by Sanger sequencing (section 2.7) with the appropriate primer. Confirmed clones were expanded to maxi-cultures, with the addition of 1 mL from mini-culture in 500 mL of LB with the appropriated antibiotic selection and cultured overnight at 37 °C in shacking incubator. Plasmid extraction was performed using the NucleoBond Maxi Kit (*Macherey-Nagel GmbH*), as described in section 2.4.2.

2.17.6 *Cryopreservation of Bacterial clones*

Bacterial clones from LB liquid cultures were mixed with an equal volume of 40% glycerol in cryovials and stored at -80 °C. The 40% glycerol stock was sterilised by autoclaving.

2.18 Mammalian cultures

2.19 Cells lines and Culture conditions

Table 8: Mammalian cell lines

Cell line	Origin	Ref.
HEL	Human erythroid leukaemia cells	[229]
MEL	Murine pro-erythroid leukaemia cells	[230]
HEK 293T	Human embryonic kidney cells	[231]

2.19.1 *HEK 293 cells*

HEK 293T cells were cultured in complete (c) Iscove's Modified Dulbecco' Medium (IMDM) (*Invitrogen™, Thermo Fisher Scientific*) medium (i.e., including 10% FBS, 1% penicillin/streptomycin and 1% 100-mM L-Glutamine, (all *Invitrogen™, Thermo Fisher Scientific*), in 75-cm^2 flasks or 15-cm diameter dishes (*Corning, NY, USA)* and incubated at 37 °C in a 5% CO_2 atmosphere in a NuaireDH Auto Flow CO_2 Air Jacket incubator (*Plymouth, USA*). Medium was replaced every 2–3 days, and cells were passaged when they reached 70–80% confluency.

2.19.2 *Human erythroleukaemia (HEL) cells*

HEL cells were cultured in cRPMI (supplemented as before) in 10-cm diameter dishes (*Corning, NY, USA*) and maintained at 250,000–500,000 cells/mL.

2.19.3 *Murine erythroleukaemia (MEL) cells*

MEL cells were cultured in cRPMI (supplemented as before) in 10-cm diameter dishes (*Corning, NY, USA*) and maintained at 250,000–500,000 cells/mL. Terminal erythroid differentiation of these cells was induced by addition of 1.5% DMSO to cultures of 100–200 000 cells/mL. Cells were collected for RNA and protein analysis at day 3 and day 6 and 9 post-differentiation, respectively.

2.19.4 *Cryostorage of mammalian cell lines*

Cell lines, HEK293T, HEL and MEL, were cryopreserved at -80 °C in freezing medium (20% FBS (*Invitrogen™, Thermo Fisher Scientific*), 10% DMSO (*Sigma-Aldrich, UK*) and 70% respective culture medium), using a controlled-rate freezing container (*Nalgene, NY, USA*). Frozen cells were then transferred in liquid nitrogen for longer term storage.

2.19.5 *Thawing of mammalian cells lines*

Cells were quick-thawed at 37 °C and washed once in 10 volumes of complete respective medium to remove residual DMSO. Cells were collected by centrifugation at 300 x g for 5 minutes and pellets re-suspended in the appropriate medium for expansion.

2.19.6 *Plasmid DNA delivery methods*

2.19.6.1 *PEI transfection (HEK293T) - Targeted disruption assessment*

At 12 hours prior to polyethylenimine (PEI) (*Polysciences, Inc., USA*) mediated transfection, 150 000 HEK293T cells were seeded per well in a 24-well plate in 450 µL cRPMI (10% foetal bovine serum, 1x penicillin / streptomycin (*Invitrogen™, Thermo Fisher Scientific*), 1x GlutaMax (*Invitrogen™, Thermo Fisher Scientific*). HEK293T cells were transfected with a total plasmid amount of 1.25 µg from which 1.15 µg were constructs for specific designer nucleases, 600 ng for each TALEN monomer construct (pVAX_TALEN_FokI) (1:1 ratio) or 400 ng gRNA (pLML3636_specific-gRNA) and 800 ng for Cas9 (pJDS246) (1:3), and 50 ng of pRK5.mCherry transfection reporter plasmid [228].

When targeted disruption was measured by flow cytometry, 100 ng of $HBB^{IVSI-110}$_GFP reporter plasmid (section 2.8.3) and 100 ng of pLSS mOrange N1 (*Addgene, UK*) reporter plasmid were used for the assessment of targeted disruption and transfection efficiency, respectively. Designer nuclease constructs were adjusted appropriately keeping the ratios constant, whereas pUC118 plasmid was used to normalize the amount of plasmid DNA (e.g., for designer-nuclease-negative control samples). The plasmid DNA mixture was made up to 25 µL with 150 mM NaCl (pH 5.5). Then, 25 µL of PEI mix (0.1 mg/mL PEI, 150 mM NaCl pH 5.5) was added to the plasmid DNA/NaCl mixture, mixed and incubated for 10 minutes at RTemp. Then, PEI / plasmid DNA / NaCl mixture was added drop-wise to the cells. Transfected cell were cultured for 12 hr at 37 °C, before changing the medium. Cells were detached by treatment with trypsin and harvested 48 hr post-transfection and used for the assessment of transfection (pEGFP N1 or pLSS mOrange N1), targeted disruption ($HBB^{IVSI-110}$ eGFP reporter) efficiencies by flow cytometry and gDNA extraction (*Qiagen Blood Mini Kit, Qiagen, Hilden, Germany*).

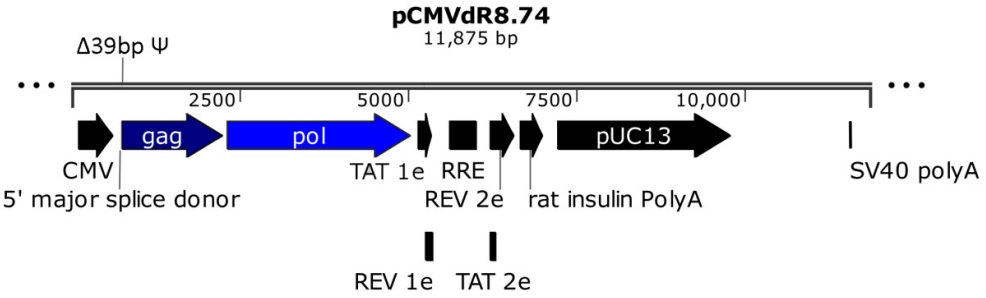
2.19.6.2 *Electroporation of transgenic murine erythroleukaemia cells*

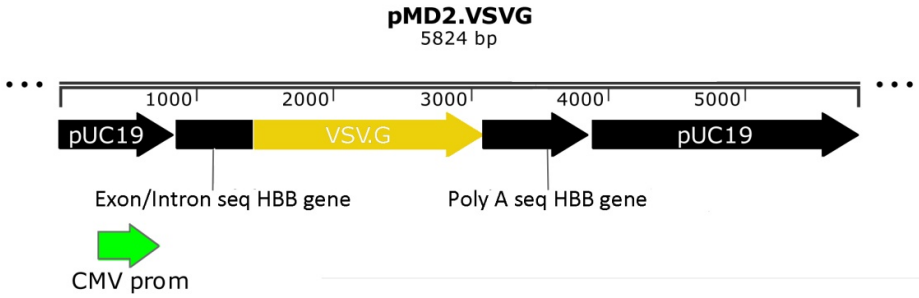
$HBB^{IVSI-110}$ - specific designer nucleases were delivered to MEL MA821 $HBB^{IVSI-110}$ cells by electroporation. Cells were washed once in 40 mL RPMI medium and centrifugation (200 x g 10 minutes). Cell pellets were resuspended at a cell density of 10 x 10^6 cells/450 µL RPMI medium and 60 µg total plasmid DNA (in 50µl of RPMI medium) was added followed by electroporation (500 µL V_f

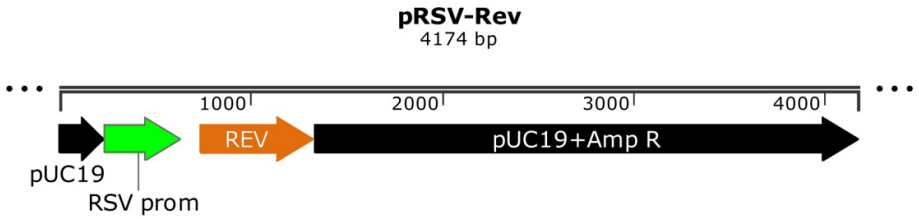
/ 400 V / 850 µF / Infinite Resistance). $HBB^{IVSI-110}$ specific designer nuclease plasmids constituted 2/3 of the total plasmid amount (40 µg) keeping the appropriate ratios constant (TALEN monomers 1:1, RGEN 1:3), whereas the remaining 1/3 was pEGFP_N1 (*Clontech, USA*) plasmid from now on reffered as CMV GFP reporter (20 µg). Normalization of the amount of plasmid DNA was by addition of pUC118 plasmid. Electroporated cells were cultured in 9 mL cRPMI for 6 hr at 37 °C (recovery period) before being moved to hypothermic conditions at 30 °C for 72hr. Cells were then collected for assessment of transfection efficiency (by flow cytometry) and gDNA extraction for targeted disruption assessment (T7E1 Assay / TIDE), while cultures were further expanded at 37 °C for 3–4 days before proceeding in the induction of erythroid differentiation. Genome edited MEL MA821 $HBB^{IVSI-110}$ cells (100 000 cells/mL) were induced to undergo terminal erythroid differentiation in 10mL MEL differentiation medium (cRPMI supplemented with 1.5% DMSO) for 9 days. Induced cells (5 mL) were collected on days 3 and 6 for RNA extraction ($5x10^6$ cells/1mL Trizol) and day 6 for protein extraction ($5 x 10^6$ cells/50 µL RIPA lysis Buffer).

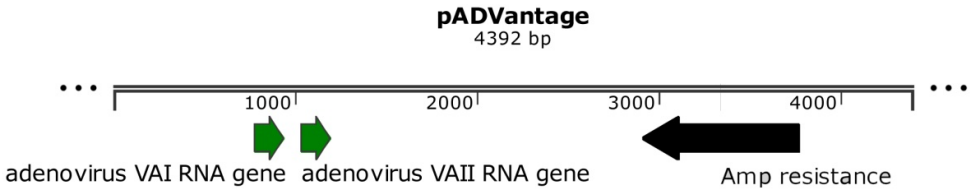
2.20 Lentiviral vector production (calcium phosphate transfection)

Vectors were produced by adoption of published methods [232]. Briefly, 24 hours prior to transfection $9x10^6$ HEK 293 T cells were seeded into 15-cm diameter tissue culture plates with 20 mL cIMDM. At two hours before transfection the medium was changed to 22 mL fresh cIMDM. For each plate a plasmid DNA mixture was prepared in 50 mL tubes with 9 µg of pMD2.VSVG (envelope plasmid), 16.25 µg of pCMVΔR8.74 (2^{nd} generation of gag/pol packaging construct), 6.25 µg of pRSV-REV (3^{rd} generation REV construct), 15 µg of pADvantage (encodes the adenovirus associated RNA genes, VAI and VAII, which can enhance transient protein expression in a variety of cell types by increasing translation initiation [233]; (Figure 18) and 30 µg transfer vector (Table 24). The DNA mixture was made up to 1125 µL with 0.1X TE and after addition of 125 µL of 2.5 M $CaCl_2$ (*Sigma-Aldrich*) was incubated for 5 minutes at RTemp. While vortexing at full speed, plasmids were precipitated by dropwise addition of 1250 µL 2x HBS solution (281 mM NaCl (*Sigma-Aldrich*), 100 mM HEPES (*Scharlau*), 1.5 mM Na_2HPO_4 (*Sigma-Aldrich*), pH 7.12) to the 1250-µL mixture, immediately before the precipitate solution was added to the HEK293T culture and incubated for 16 hr at 37 °C and a subsequent medium change with 16 mL cIMDM supplemented with 1-mM sodium butyrate (*Sigma-Aldrich, UK*). Vector-containing supernatant was collected 48 h after transfection.


pCMVdR8.74
11,875 bp
Δ39bp Ψ
2500
5000
7500
10,000
gag
pol
pUC13
CMV
5' major splice donor
TAT 1e
RRE
REV 2e
rat insulin PolyA
SV40 polyA
REV 1e
TAT 2e


pMD2.VSVG
5824 bp
1000
2000
3000
4000
5000
pUC19
VSV.G
pUC19
Exon/Intron seq HBB gene
Poly A seq HBB gene
CMV prom


pRSV-Rev
4174 bp
1000
2000
3000
4000
REV
pUC19+Amp R
pUC19
RSV prom


pADVantage
4392 bp
1000
2000
3000
4000
adenovirus VAI RNA gene
adenovirus VAII RNA gene
Amp resistance

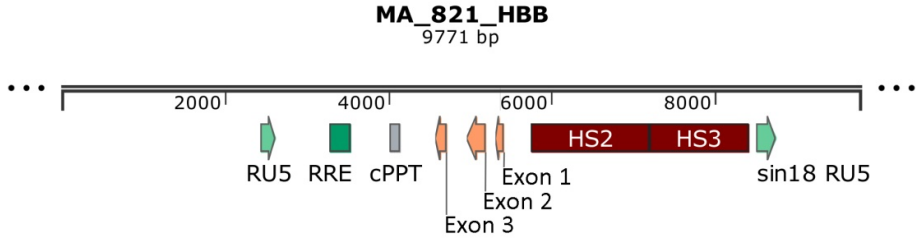

MA_821_HBB
9771 bp
2000
4000
6000
8000
HS2
HS3
RU5
RRE
cPPT
Exon 1
Exon 2
Exon 3
sin18 RU5

Figure 18: Constructs used for lentiviral vector production.
pMD2.VSVG: envelope constructs which encodes heterologous vesicular stomatitis virus G glycoprotein (VSVG) for LV pseudotyping driven by the CMV promoter and SV40 poly A. **PCMΔ874 construct**: 2nd generation packaging plasmid which encodes the gag gene for capsid proteins; the pol gene for the viral enzymes, protease, integrase, RNase H and reverse transcriptase essential for viral replication, Rev gene for rev proteins and Tat for trans-activating regulatory protein. **pRSV-REV**: 3rd- generation construct encoding the REV protein under the RSV promoter. **pADvantage**: Construct encoding the adenovirus RNA I and II genes. Enhances transient protein translation in a variety of cell types by increasing the initiation of translation. Example of **Transfer vector** : **MA821 HBB** ("GLOBE") harbouring *HBB* transgene [132]. *HBB* Exons 1 – 3. CMV: cytomegalovirus promoter. RSV: Rous sarcoma virus promoter. RRE: REV responsive element. cPPT: central polypurine tract. HS2, HS3: DNase I hypersensitive elements 2 and 3 of the *HBB* locus control region. RU5: LTR element. sin-18 RU5: self-inactivated 3' LTR. Δ39 bp Ψ packaging signal: inactivated Ψ signal with 39bp deletion. Amp: ampicillin resistance.

2.20.1 *Concentration of LVs*

LV-containing supernatant was centrifuged for 5 minutes at 200 x g and the resultant supernatant passed through a Millex-HV 0.45-*μ*m pore size PVDF membrane filter (*Millipore, Darmstadt, Germany*) to remove any cells or debris. LVs were then concentrated from this supernatant via centrifugation at 20,000 x g for 4 hr at 4 °C, the pellet gently re-suspended in sterile PBS (*Gibco/Life Technologies*, 1/350 of the starting volume) and the tube rotated for 1 hr at 4 °C before storage of aliquots at -80 °C in cryotubes (*Nunc; Thermo Fisher Scientific*).

2.20.2 *Titration of LVs*

At 24 hr prior to transduction, $2x10^5$ HEL cells were seeded at a concentration of $2.0x10^5$ cells/900 µL on 24-well plates in cRPMI with the volumes as shown in Figure 19 (1st well 990 µL and the rest 900 µL). On the day of the transduction, before the addition of the LVs, polybrene (*Sigma-Aldrich*) was added to all the wells at a final concentration of 8 µg/mL. Cells in the 1st well were transduced with 10 µL of concentrated LV mixed well (10^{-2} dilution) and then 100 µL of cells/LVs mixture was transferred to the next well (10^{-3} dilution), mixed again and move to the next until the last well (10^{-7} dilution) in which the 100 µL are discarded. In this way cells were transduced with 10-fold serial dilutions (10^{-2}–10^{-7}) of the concentrated LV stock whilst maintaining the final volume, the number of transduced cells and polybrene concentration constant. One well was left untransduced as a negative control. Biological titres were determined by quantitative PCR (qPCR) for all vectors.

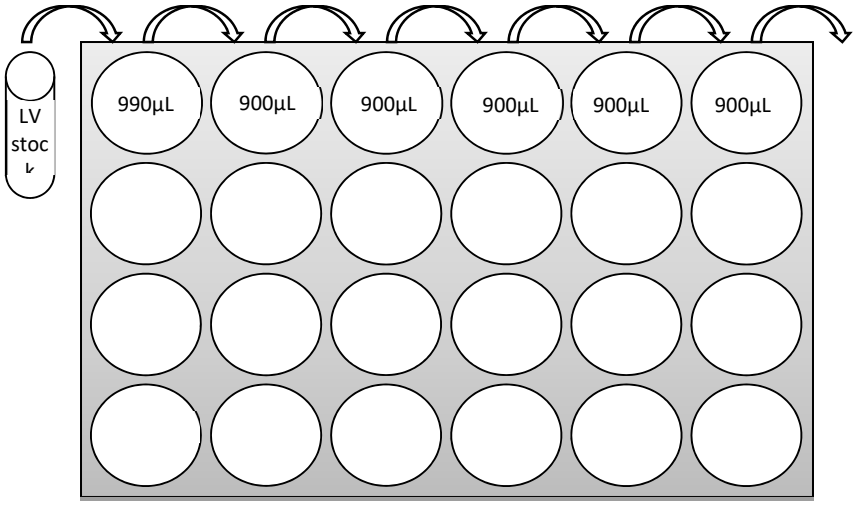


Figure 19: Set up of LV titration plate.
At 24 hr prior to transduction, $2x10^5$ HEL cells were seeded at cell concentration of $2.0x10^5$ cells/900 µL on 24-well plates in cRPMI at the volumes shown (1st well 990 µL and the rest 900 µL). On the day of the transduction, before the addition of the LVs, polybrene was added to all the wells at a final concentration of 8 µg/mL in all the wells. Cells in the 1st well were transduced with 10 µL of concentrated LV mixed well and then 100 µL of cells/LVs mixture was transferred to the next well, mixed again and moved to the next well maintaining the final volume, transduced cell number and polybrene concentration constant. Untransduced cells were also maintained as negative control.

2.20.3 ***Biological titres via qPCR***

2.20.3.1 *Vector copy number quantification: Single-plex qPCR*

Titres of all the produced LVs are measured by qPCR at 20 days after transduction by analysing the gDNA of transduced cells on a 7900HT Fast Real Time PCR System with SDS version 2.2.2 software (*Applied Biosystems, MA, USA*) and the default cycle conditions: 2 minutes at 50 °C, 10 minutes at 95 °C, (15 seconds at 95 °C, 1 minute at 60 °C) x 40 cycles. The qPCR reactions were prepared in 96-well plates. Standard curves for gDNA (200, 50, 5, 0.5 ng) and LV plasmid equivalents of VCN 20, 4, 0.8 and 0.16 in 100 ng of human gDNA were generated for absolute quantification of LV titres in 25 µL reactions. The MA821 HBB^{T87Q} LV construct (9771 bp) vector encoding the transfer vector segment with the HBB^{T87Q} transgene, was used for the preparation of the VCN plasmid standard curve following ABI guidelines [234] in which 3.26 pg of plasmid is equivalent to VCN 1 in 100 ng of 3.1 Gb human haploid genome. The highly conserved poly (rC) – binding protein 2 (PCBP2) endogene and LV LTR region were used as target sequences for gDNA and LV detection respectively, using the primer and probe sequences listed in Table 22. Briefly, singleplex qPCR reactions consisted of 12.5 µL of 2x TaqMan PCR Master mix (*Applied Biosystems, Thermo Fisher Scientific)*, 900 nM of each primer, Lenti or PCBP2 pair, 250nM of specific probe (Lenti VIC probe or PCB2_FAM probe), 2 µL of plasmid DNA SC dilutions or 100ng-of sample gDNA (50 ng/µL) and made up with water to 25 µL final volume. Real-time qPCR reactions started with activation of the polymerase at 95 °C for 15 minutes, followed by 40 cycles of denaturation at 95 °C for 30 seconds and annealing / extension at 60 °C for 1 minute, respectively. Based on the Lenti plasmid- and PCBP2 gDNA- standard curves, absolute quantities of each samples were quantified leading to the assessment vector copy number (VCN) and titres.

Normalised VCN (a) and Titres (b) were calculated using the following formulas:

a) Normalized VCN = VCN (Lenti Quant.) x 100 ng theoretical gDNA /gDNA (PCBP2 Quant)

b) Titres (Transduction units (TU)/mL) = (Normalized VCN)*Vector dilution*(total no. of cells)

2.20.3.2 *Vector copy number quantification: Multiplex qPCR*

A new multiplex qPCR protocol for VCN quantification using the Qiagen Multiplex PCR kit, was developed and used for the quantification of VCN in transduced MEL MA821 $HBB^{IVSI-110}$ VCN1 and human $CD34^{+}$ cells, as described [235]. MA821 HBB^{T87Q} LV construct (9771 bp) vector encoding the transfer vector segment with the HBB^{T87Q} transgene, was used for the preparation of the VCN plasmid standard curve following ABI guidelines [234] in which 3.26 pg and 3.74 pg of plasmid is equivalent to VCN 1 in 100 ng of human (3.1 Gb/haploid genome) and murine (2.7Gb/haploid

genome). For the LV standard curve, plasmid equivalents of VCN 20, 4, 0.8 and 0.16, in 100 ng of murine or human gDNA and spiked with 100 ng gDNA were generated for absolute quantification of the VCN of transduced MEL or $hCD34^+$ cells in 25 µL reactions. Murine or human gDNA was used for the preparation of the gDNA standard curve at dilutions of 200, 100 and 50 ng per 25 µL reaction spiked with plasmid equivalent with VCN 0.8 in 100 ng of genome DNA. Default cycle conditions were used as described above and sequences of Lenti and PCBP2 specific primer pairs and probes (-Multi) are listed in Table 22. Briefly, multiplex qPCR reaction is consisted with 12.5 µL of 2x Multiplex Qiagen master mix, 200µM dNTPs, 3.5 mM $MgCl_2$, 300 nM of each primer pair, qLenti and qPCBP2, 250nM of specific probes (qLenti VIC probe and qPCBP2_FAM probe) and as DNA template 2 µL of VCN or gDNA SC dilutions or 100 ng of sample gDNA (50 ng/µL) and made up with water to 25 µL final volume. Standard PCR conditions used starting with 2min at 50 °C, activation of the polymerase at 95 °C for 10 minutes, followed by 40 cycles of denaturation at 95 °C for 15 seconds and annealing / extension at 60 °C for 1 minute, respectively. Then, VCN was measured: initially automatic outlier removal for replicate samples was performed (above 1.5x standard deviations from the triplicate median was set as a cut – off value), by calculating the paired Student's t-test for PCBP2-mediated detection of the invariable gDNA amount (either 50, 100 or 200 ng), and by calculating the squared Pearson's correlation coefficient (R^2) between true VCN values (as defined by manually added plasmid amounts) and ΔΔCt values for LV-mediated detection of LV serial dilutions (0.16, 0.8, 4, 20). To this end and excluding the use of an additional calibrator gene, ΔCt was defined as 0 for all standard curve samples. Within each run, reproducibility of C_t values for invariable PCBP2 template amounts was assessed as a scaled population-wide standard deviation, and correlation of ΔΔCt values for LV dilutions as the R^2.

2.21 LV Transduction of MEL cells

On the day of transduction, 2 x 10^5 MEL cells were placed in 2 mL tubes in a final volume of 0.5-1 mL cRPMI supplemented with 8 µg/mL polybrene. MEL cells were transduced with different MOIs with pLKO.I shIVS1-110 (Up/Mid /Mid2/Down) and shScramble LVs. One sample was left untransduced as a negative control. The transduction procedure includes mixing of cells/LV mixture by pipetting 2–3 times every hour for 6 hours. Transduced cells were centrifuged at 300 x g for 5 minutes and cell pellets were re-suspended in 1 mL cRPMI and seeded in 24-well plate for expansion. VCN was assessed using the quantitative PCR (qPCR) for all vectors as described in section2.20.3.2

2.22 Production of Transgenic MEL cell lines

2.22.1 *MEL MA821$^{IVS1-110}$ and MA821 cell pool (average VCN 1.9 - 2)*

MEL cells were transduced with a range of MOIs with LVs MA821 HBB$^{IVS1-110}$ and MA821 HBB^{Normal} following the methodology described in section 2.22. Cells were cultured for ~2 weeks before VCN was assessed by qPCR (section 2.20.3.2). Transgenic MEL cells for MA821$^{IVSI-110}$ and MA821 HBB^{Normal}, with average VCN 1.9 and 2, respectively, were selected as cell lines for pool cell experiments.

2.22.2 *Clonal selection of transgenic MEL MA821$^{IVSI-110}$ clones with known VCN (1,2,3,4)*

Initially, transgenic MEL MA821$^{IVSI-110}$ (average VCN 1.9) cells were diluted to working concentrations of 1.5 cells / 100 µL cRPMI and seeded into six 96-well plates. Briefly, working concentration of 50 000 cells/mL was prepared followed by 10-fold serial dilutions to the theoretical working concentration of 1.5 cells/100 µL in a final volume of 50 mL. A 100 µL aliquot of cell suspension was seeded per well of a 96-well plate. Cells were cultured for 48hr without moving the plates to achieve distinguishable growing colonies that were easy to score under the microscope. Wells with single clones were expanded to 24-well plates. VCN was assessed by qPCR as described in section 2.20.3.2 and clones with VCN 1, 2, 3 and 4 were cryopreserved at -80 °C as described in section 2.19.4.

2.23 Clonal selection of genome edited MEL MA821$^{IVSI-110}$ VCN 1

MEL MA821 HBB$^{IVSI-110}$ VCN 1 cells genome edited with HBB$^{IVSI-110}$ specific designer nucleases (HBB TAL R1/L1, R1/L2 and HBB$^{IVSI-110}$ RGEN) were diluted to a cell suspension concentration of 1 cell / 200 µL cRPMI medium in a final volume of 50 mL and 100 mL for TALEN's and RGEN's edited cells, respectively. All cells were seeded in 96-well plates with 200 µL/well. Plates were incubated for 48 hours before checking under the microscope for single colonies. Wells with single clones were expanded to 24 well plates to reach to an appropriate number of cells that allowed gDNA extraction. All treated clones with HBB TAL R1/L1 (69), HBB TAL R1/L2 (83), HBB$^{IVSI-110}$ RGEN (143) and pUC118 (24) were cryopreserved as described in section 2.19.4 until genome edited clones were characterized by Sanger sequencing (section 2.23.2).

2.23.1 *Characterization of INDELs*

For the characterisation of INDELs produced by the NHEJ repair pathway after treatment of MEL MA821$^{IVS1-110}$ cell pool with specific designer nucleases HBB TAL R1/L1 and R1/L2, HBB$^{IVSI-110}$ RGEN and nuclease-free negative control pUC118, PCR products encompassing the target site were cloned into pCR.4 Blunt-TOPO vectors using Zero Blunt® TOPO® PCR Cloning Kit (*Invitrogen^{TM}, Thermo Fisher*

Scientific) for sequencing, according to manufacturer's instructions. The pCR 4 Blunt-TOPO vector is provided linearized, with 3' end of each strand to be covalently bound to Vaccinia virus DNA topoisomerase I, which allows the efficient cloning the PCR products in the vector. Briefly, PCR products were produced in a (50 µL) PCR reaction, using Phusion polymerase (section 2.9.2), which creates blunt-ended PCR products, the primer pair (GCB1 FW and 109 RV) that encompasses Exon 1, Intron 1 and part of Exon 2 and as template 200 ng of gDNA from the genome edited MEL MA821$^{IVSI-110}$ cell pool. The TOPO cloning reaction was performed with the use of 3 µL of fresh PCR product, 1 µL of TOPO salt solution, 1 µL of TOPO vector and made up with water to a final volume of 6 µL. The reaction mixture was incubated for 20 minutes at RTemp. Then, 2 µL of the TOPO cloning reaction was used for the transformation of the TOP10 chemically competent bacteria (section 2.17.3), from which positive clones were selected on LB agars with 50 µg/mL of kanamycin. A total of 100 colonies were picked for the characterisation of INDELs produced by designer specific nucleases, HBB TAL R1/L1, HBB TAL R1/L2 HBB$^{IVSI-110}$ RGEN and ~30 colonies from nuclease-free negative control, pUC118. Colonies were cultured in 5 mL LB (50 µg/mL kanamycin) in shacking cultures overnight at 37 °C, 0.5% CO_2. Bacterial plasmid DNA was extracted (section 2.4.1) and sequenced using M13 FW primer (section 2.7). Alignment of the sequencing traces was performed using SnapGene software (GSL Biotech; available at snapgene.com).

2.23.2 *Characterization of genome-edited MEL MA821 HBB$^{IVSI-110}$ clones*

A genome edited population of the MEL MA821 HBB$^{IVSI-110}$ VCN 1 cell clone was seeded in a 96-well plate for selection of edited clones. Clones were expanded to 24-well plates and gDNA extracted using the FlexiGene kit (*Qiagen*) according to manufacturer's instruction. A 100 ng aliquot of gDNA was used as template for a Phusion PCR reaction with specific primers encompassing the target sequence (GCB1 FW /109 RV, Table 22) in a 25ul final volume reaction. PCR products were treated with the ExoSap kit (*Affymetrix, Thermo Fisher Scientific*) according to manufacturer's instructions (section 2.10.2); ExoSap treated products were then sequenced using HBB Ex2.2 RV primers. Genome edited clones were cultured and differentiated with cRPMI supplemented with 1.5% DMSO. Induced cells were collected (5 mL) on days 3, 6 and 9 for RNA extraction (5x10^6 cells/1mL TriFast) and protein extraction (10^6 cells/50 µL RIPA lysis Buffer).

2.24 Human CD34$^+$-derived erythroid progenitor cell liquid cultures

2.24.1 *Culture conditions*

2.24.1.1 *Isolation of mononuclear cells from peripheral blood*

Peripheral blood (PB) from volunteers was collected in EDTA-coated tubes (*Greiner Bio-One; Frickenhausen, Germany*). PB was mixed with 1.5 volumes of PBS (without $CaCl_2$/$MgCl_2$) to reduce erythrocyte clumping in 50 mL times. Then, 1.25 volumes of RTemp density gradient reagent (*Lymphoprep, Accu-Prep^{TM} Lymphocytes*, *Axis-Shield PoC AS, Norway*), was added in 50 mL tubes and keeping the tube at 45° the PB-PBS mixture was carefully overlaid on top of the Lymphoprep reagent creating two phases. Mononuclear cells were then separated via centrifugation at 900 x g, w/o brake, for 25 minutes. The interphase, which consists of nucleated cells was collected with a 5 mL pipet with caution not to take Lymphoprep from the lower phase. Mononuclear cells were washed 3 times in PBS, twice by centrifugation at 300 x g for 5 minutes and once at 200 x g for 5 minutes, all at RTemp. During washes the cell pellet was loosened by flicking the tube 3–4 times. The final washed cell pellet was re-suspended in 600 µL ice-cold 1% bovine serum albumin (BSA) in PBS.

2.24.1.2 *Isolation of human CD34$^+$ cells*

Peripheral blood mononuclear cells (PBMC) suspended in 600 µL ice-cold 1% bovine serum albumin (BSA) in PBS were incubated for 15 minutes with 100 µL anti-hCD34$^+$ MicroBeads (*Miltenyi Biotecy, Bergisch Gladbach, Germany*) on ice on a rocking platform at 4 °C. Cells were washed twice with 14 mL ice cold beading buffer (BB) (0.5% BSA, 2 mM EDTA in PBS (without $CaCl_2$/$MgCl_2$) and centrifugation at 300 x g for 5 minutes at 4 °C with brake on and maximum acceleration. Cell pellets were re-suspended in 1 mL BB (EDTA-free). Magnetic bead separation was carried out with LS MACS columns (2 columns / sample) (*Miltenyi Biotecy*). The MidiMACS separator and Multistand (*Miltenyi Biotecy*) were pre-chilled by incubation for ≥3 hours at 4 °C. LS MACS columns were placed on the magnetic adaptor and equilibrated with 5 mL BB. Cells were loaded onto the column and for every 0.5 mL of cells the column was washed with 1.5 mL BB. The column was washed twice with 3 mL BB. Cells were eluted from the column by removing it from the magnetic separator and immediately adding 5 mL BB. Cells were then flushed into a new 15-mL tube by inserting the plunger into the column. Eluted cells pass from a second round of selection onto a new equilibrated column by loading 2.5 mL of cells and washing with 2.5 mL BB following the same procedure as with the first selection. Cells were eluted in 5 mL BB. Cells were collected by centrifugation at 300 x g for 5 minutes at RTemp. The CD34$^+$ enriched cell pellet was re-suspended in 3 mL expansion medium (see

section 2.24.1.3) in 6-well plates (*Nunc; Thermo Fisher Scientific*) and cultured under default conditions (37 °C, 5% CO_2 atmosphere).

2.24.1.3 *hCD34$^+$ culture*

The hCD34$^+$ cells after isolation through MACS columns were re-suspended in 3 mL expansion medium consisting of StemSpan II medium (*Stemcell Technologies, Vancouver, Canada*) supplemented with 1x CC100 (*Stemcell Technologies*), 2 unit/mL erythropoietin (*Sandoz GmbH, Austria, Binocrit; 4,000 IU/0.4 mL*), 10^{-6}M dexamethasome (*Sigma-Aldrich*) and 1x Penicillin/Streptomycin (*Corning-CellGro*) and seeded into 6-well plates (*Nunc; Thermo Fisher Scientific*). Cells were cultured for 72 hr before completely changing the medium after centrifugation at 300 x g for 5 minutes. Then, on day 6 following expansion cells were transferred to 25 cm^2 Nunclon^{TM} Delta flasks (*Nunc; Thermo Fisher Scientific,*) with 7 mL expansion medium. On day 10 of expansion cell number should reach from 10 to several millions of cells. In general, cell densities should be kept below 0.5 x 10^6 cells/mL in the medium specified above and medium should be changed twice every week based on confluency.

Following electroporation hCD34$^+$ cells were cultured in 6 mL expansion medium without antibiotics in 25 cm^2 Nunclon^{TM} Delta flasks. Antibiotics are avoided in the expansion medium of electroporated cells as this is associated with increased cell death.

2.24.2 ***Freezing and thawing hCD34$^+$ cells***

Human CD34$^+$ cells were frozen at a minimum concentration of 0.5 x 10^6 cells/mL in ice cold freezing medium, 50% characterized FBS (*Hyclone, Thermo Fisher Scientific*), 10% DMSO in IMDM medium. Initially, enriched hCD34$^+$ cell cultures following expansion for 7–10 days were counted, pelleted by centrifugation at 300 x g for 5 minutes, re-suspended in the appropriate volume of ice cold freezing medium (1–4 x 10^6 cells/mL) and 1 mL aliquots were prepared in ice-cold cryovials. Cryovials were then transferred to a controlled-rate freezing isopropyl-containing, Mr. Frosty container and stored at -80 °C. Cryovials were stored for up to 30 days at -80 °C, and in liquid nitrogen for longer term storage.

2.24.3 ***Thawing hCD34$^+$ cells***

Frozen hCD34$^+$ cells were fast-thawed by incubation at 37 °C and transferred drop-wise into 10x volume of thawing medium (IMDM medium supplemented with 5% characterized FBS). Cells were incubated for 10–15 minutes at RTemp and pelleted by centrifugation at 200 x g for 10 minutes with acceleration 6 and full brake. Cells were re-suspended in expansion medium at 0.5–1 x 10^6 cells/mL.

2.24.4 *Induced differentiation of hCD34$^+$ cells*

Erythroid differentiation of hCD34$^+$ cells was induced by culturing in differentiation medium consisting of 70% α-minimum essential medium (αMEM) (*Corning CellGro*; 10-22-CV), 30% FBS (*Hyclone*), 10^{-5}M 2-mercaptoethanol (*Sigma-Aldrich*), 10 units erythropoietin (*Binocrit 4,000; Sandoz GmbH*), 10ng/mL stem cell factor (SCF) (*Peprotech*) and 1% Penicillin / Streptomycin (*Invitrogen™, Thermo Fisher Scientific*). Cells were differentiated at a concentration of 0.5–1 x 10^6 cells/mL for 6 and 7 days in the case of genome edited and transduced hCD34$^+$ cells, respectively at 37 °C in a 5% CO_2 humidified atmosphere.

2.24.5 *Cytocentrifugation (Dianisidine / May-Grünwald / Giemsa staining)*

Phenotypic characterization of erythroid progenitors in primary cultures of differentiated hCD34$^+$ cells was based on morphology using cytocentrifugation preparations. Cell suspensions of 0.5–1 x 10^4 hCD34$^+$ cells in 200 µL PBS were prepared. A 100–200 µL aliquot of cells was loaded per cytocentrifugation chamber and transferred onto the slides by centrifugation at 95 x g for 3 minutes (acceleration: 6, break: 6) using the Tharmac Cellspin II cytocentrifuge with an EASY rotor (*Hettich, Germany; A320*). Cell smears were air-dried and fixed onto the slides for 10 minutes by treatment with in cold methanol followed by 2 minutes in dianisidine solution (1.5% o-Dianisidine (*Sigma-Aldrich*) in methanol), 2 minutes in H_2O_2/ethanol solution (50% EtOH, 1.5% H_2O_2 in ddH_2O), rinsed with ddH_2O, 5 minutes at RTemp in May-Grünwald staining solution (*Sigma-Aldrich*), rinsed with ddH2O, 10 minutes at RTemp in Giemsa's staining solution (*Fluka Analytica*, *Sigma-Aldrich*), rinsed with ddH_2O and then allowed to air dry. Finally, cell spots were covered with mounting medium (*DakoCytomation, CA, USA*) and sealed under a coverslip. Slides were viewed under the IX73 inverted microscope (Olympus Corporation) equipped with a camera for capturing images.

2.24.6 *LV transduction hCD34$^+$ cells*

LV transduction was conducted with 1x10^6 hCD34$^+$ cells in 0.5 mL expansion medium supplemented with 8 µg/µL polybrene (*Millipore*) in 2 mL tubes. LV stocks were thawed on ice and used to transduce cells with a range of MOIs to give similar VCN. After addition of LV t cells were incubated for 6 hour at 37 °C in a 5% CO_2 humidified atmosphere, in which cells/LVs were mixed by pipetting every hour. After six hours, cells were centrifuged at 200 x g for 5 minutes and re-suspended in 3 mL expansion medium. Cells were collected 48 hr post-transduction for further analysis (gDNA extraction for VCN quantification; Dianisidine, Giemsa and May-Grünwald staining; 2.24.5) and differentiation. A 1–2 x 10^6 aliquot of cells were cultured in 3 mL differentiation medium. Cells were collected on day 6 post-differentiation for protein extraction and HPLC analysis. Cell death, level of

differentiation (Hb expression) and morphology were assessed microscopically by trypan blue and dianizidine and May-Grünwald-Giemsa staining, respectively.

2.24.7 *Electroporation of human CD34$^+$ cells*

Cryopreserved hCD34$^+$ cells were fast-thawed at 37 °C and transferred drop-wised to IMDM medium supplemented with 5% characterized FBS. Cells were incubated for 10 minutes at RTemp and centrifuged at 200 x g (9 acc/6 dec) and pellets re-suspended in 9mL expansion medium. Medium was changed one day prior to electroporation (day 3). Cells were washed once in 8 mL StemSpam II medium (200 x g, 10 min). Cell pellets were diluted to a cell density of ~5 x 10^6 cells / 450 µL StemSpam II to which was added 40 µg of total plasmid DNA in 50 µL StemSpam II medium and the mixture subjected to electroporation using a Bio-Rad Gene Pulser unit (500 µL V_f 300 V / 1050 µF / Infinite Resistance). HBB$^{IVSI-110}$ specific designer nuclease plasmids constituted 7/8 of the total plasmid amount (35 µg) keeping the appropriate ratios constant (TALEN monomers 1:1, RGEN 1:3), whereas the remaining 1/8 was pEGFP N1 reporter plasmid (5 µg). Addition of pUC118 plasmid was used to normalize for the amount of plasmid DNA. Electroporated cells were cultured in 6 mL expansion medium without antibiotics (penicillin, streptomycin) in 75 cm^2 Nunclon^{TM} Delta flasks for 48 hr at 37 °C in 5% CO_2 humidified atmosphere. Antibiotics are avoided in the expansion medium of electroporated cells as this is associated with increased cell death. Then, cells were harvested for further analysis (gDNA extraction, 2.3.2; flow cytometry, 2.16; dianisidine / May-Grünwald / Giemsa staining; see section 2.24.5).

For erythroid differentiation, 2 x 10^6 cells were cultured in 3 mL differentiation medium. Cells were collected on day-7 for RNA (see section 2.5) and protein extraction for HPLC analysis (see section2.15). Cell viability following differentiation (Hb expression) and morphology were assessed microscopically by trypan blue and dianisidine and May-Grünwald / Giemsa staining (see section 2.24.5), respectively.

2.25 Assessing degree of genome editing (TIDE analysis)

Assessment of genome editing by designer nucleases on the target locus was measured by the online web tool TIDE (https://tide.nki.nl/)[236]. TIDE is able to quantify the editing efficacy and identify the predominant types of INDELs in the DNA of a targeted cell pool based on the quantitative sequence trace data from standard capillary sequencing reaction. Purified PCR products were Sanger sequenced (section 2.7) and traces analysed with TIDE. INDEL frequencies were measured relative to INDEL frequencies of the nuclease-free negative control sample.

2.26 In silico prediction of off-targeting potential of designer nucleases

2.26.1 *TALENS – PROGNOS web tool*

Potential off-target sites of TALEN pairs were identified by employing the PROGNOS web tool (http://bao.rice.edu/cgi-bin/prognos/prognos.cgi) using the TALEN v2.0 algorithm on the Hg19 human genome [237]. The intended binding sequence of TALEN pairs, RVD sequence of each monomer (left and right) are filled in the form. Recommended by the tool max number of mismatches per half sites were set (6), whereas as spacer length, initially the default +63 C-terminus TALEN setting were used (10–30). However, since HBB TALEN R1/L2 has an exactly 10 nt spacer length, the minimum value of the setting was reduced down to 8 in order for the programme to include the analysis of the most likely potential off target candidate, the highly homologous *HBD*, in which the spacer is 2 nt smaller compared to the *HBB* target site.

2.26.2 *RGEN - Design CRISPR web tool*

Potential off-target sites for RGEN were identified by the CRISPR design web tool (http://crispr.mit.edu/) on the Hg19 human genome [198]. Guide RNA sequence with PAM is filled on the online page. The software ranks the potential off-target sites starting from the site with the highest score for off-target binding.

2.27 Statistical analysis

Statistical analysis was conducted using Prism 7.0 (GraphPad Software Inc., CA, USA). Normality tests were used to check if data sets follow a Gaussian distribution. Prism offers three types of test based on the number of repeats in each data set, D'Agostino-Pearson (n≥8), Shapiro-Wilk (n≥7), and Kolmogorov-Smirnov (n≥5). When the *p*-value of each set is <0.05, then the null hypothesis of normal sample distribution is rejected and a non-parametric test is used. If data sets follow a normal distribution (*p*-value > 0.05), a parametric test was used for the statistical analysis of the samples (t-test, one-way ANOVA). Prism 7.0 (*GraphPad Software Inc.*) and Excel 2007 (*Microsoft, Redmond, WA, USA*) were used to draw graphs. Values are shown as ± standard deviation of the population mean.

3 Results

3.1 Development of model cell lines and quantification methods.

3.1.1 *Aims:*

- Development of an accurate multiplex RT-qPCR method for the quantification of aberrant and correctly spliced HBB mRNAs.
- Development and characterisation at transcriptional and translational level of humanised transgenic murine erythroleukaemia (MEL) cells holding the normal and mutant version of the *HBB* gene.

3.1.2 *Development of humanized MEL cell lines and methods for quantification*

Both our GT approaches for the development of personalised GT for patients carrying the $HBB^{IVSI\text{-}110(G>A)}$ mutation required initial establishment of a sensitive method for detection and accurate quantification of the normal and aberrant HBB mRNA in cells expressing the $HBB^{IVSI\text{-}110(G>A)}$ gene. The method needed to be sensitive enough to detect even small differences in the levels of both HBB mRNA variants between treated and untreated samples. Additionally, it was initially necessary to establish proof of principle in cell lines expressing the $HBB^{IVSI\text{-}110(G>A)}$ gene before moving on to assays involving patient-derived human erythroid $CD34^+$ progenitor cells. This section describes the development of a sensitive multiplex RT-qPCR, which allows the quantification of both normal and aberrantly spliced HBB transcript variants. Results are expressed as a percentage of normal and aberrant HBB mRNAs and as absolute quantities normalised to the level of erythroid differentiation. In addition, transgenic MEL cell lines were produced, which were modified to contain and express the HBB^{Normal} or mutant $HBB^{IVSI\text{-}110(G>A)}$ genes.

3.1.2.1 *Development of a novel multiplex RT-qPCR plasmid-based method for quantification of normal and aberrantly spliced HBB mRNAs.*

$HBB^{IVSI\text{-}110(G>A)}$ is a misspliced mutation that creates an abnormal SA site resulting in alternative splicing leading to an aberrant HBB mRNA containing and extra 20 nucleotides from IVS-I and the formation of a pre-mature termination codon, which are known to trigger the NMD pathway [60,238]. $HBB^{IVSI\text{-}110}$ is designated as a β^+ mutation as a small fraction of the pre-mRNA is correctly spliced, which permits the synthesis of functional HBB chains in thalassaemic$^{IVSI\text{-}110}$ patients. However, the synthesis of functional HBB chains is reduced by ~90% relative to a normal individual [90]. Quantities of the abnormally spliced mRNA were measured in previous studies by semi-quantitative methods,

such as electrophoresis gel band intensities, which limit the sensitivity and reproducibility of assessment of splicing. Expression of *HBB*$^{IVSI-110}$ in different cells lines, namely HELA and monkey kidney cells, showed that only 10–20% of the HBB pre-mRNA was correctly spliced with the rest abnormal [82,239]. A humanised β-thalassaemia mouse model harbouring a *HBB*$^{IVSI-110}$ transgene showed quantitative differences in the amount of aberrant mRNA in different tissues. In the BM and spleen the predominant mRNA was the aberrant variant, whereas in PB it was the normal mRNA [90]. Similarly, in patient reticulocytes and PB the predominant mRNA variant was the correctly spliced HBB mRNA, whereas Breda *et al*. reported the presence of the aberrant mRNA in a 1:1 ratio with the normal mRNA in liquid cultures of differentiated CD34^{+} cells from HBB$^{IVSI-110}$ thalassaemic patients [90,222,239]. In liquid cultures of mononuclear cells from thalassaemic patients the ratio of correct to aberrant mRNA was 2:1 and in some cases the aberrant mRNA was not detected [94]. These differences in the levels of the two variant HBB mRNA in different cell types are probably due to mRNA instability and the fact that HBB$^{IVSI-110}$ mRNA is subject to NMD. In addition, relative quantities were based on the semi-quantitative method of band densities of the normal and aberrant variants after RT-PCR, which is characterised by limited sensitivity.

In order to have an accurate evaluation of the therapeutic potential of both of the β-thalassaemia HBB$^{IVSI-110(G>A)}$ personalised GT approaches, a) gene addition and b) genome editing, a sensitive multiplex RT-qPCR method was developed based on variant-specific probes, a common set of primer pairs and a plasmid clone of normal and aberrant spliced HBB sequences to generate a standard curve (Figure 20). This method is a combination of absolute and relative quantification, where the final outcome is the absolute quantity of the correct and abnormal HBB mRNA molecules, which are used for the estimation of the percentage of the total HBB transcript population that is correctly or aberrantly spliced or as ratios of the correct/aberrant mRNA (ratios c/a). As described in the Materials and Methods chapter (Section 2.12.2.2.2), this method used a common pair of primers and two specific probes for the normal and aberrant mRNA (Figure 20). Different multiplex PCR kits were used during optimisation of the method, such as TaqMan Master Mix (*Applied Biosystems*) and QuantiTect multiplex kit (*Qiagen*). Ultimately, the most accurate and reproducible results were obtained with the Qiagen multiplex PCR kit (*Qiagen*).

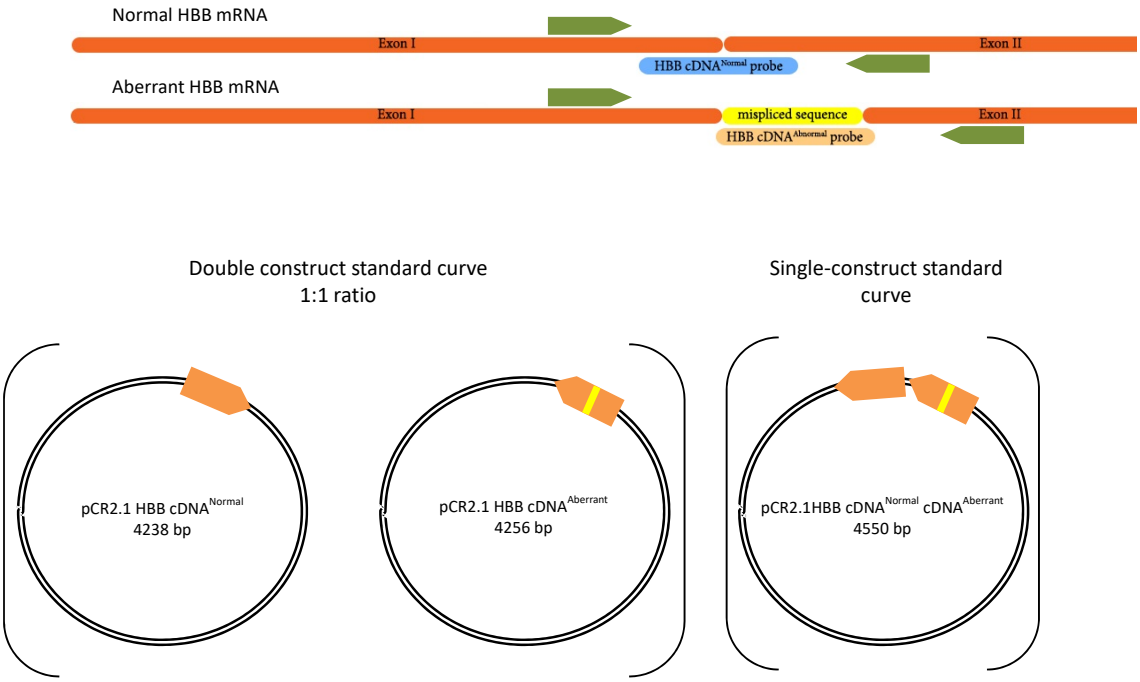


Figure 20: Primer pair, probes and plasmid constructs for multiplex RT-qPCR for the aberrant and normal HBB variants quantification.
Top: Schematic of multiplex RT-qPCR components for the quantification of the correctly and aberrantly spliced HBB mRNA. The green boxes correspond to the common pair of primers and in light blue and orange rounded boxes the correctly and aberrantly spliced HBB-specific probes respectively. Orange rounded boxes: Exon 1/2 encoded region, yellow rounded box: Abnormally spliced-in sequence (19nt) from Intron 1 in aberrant mRNA. Bottom: schematic representation of the plasmid construct used for the preparation of the double constructs standard curve (pCR2.1 HBB cDNA^{Normal} and pCR2.1 HBB cDNA^{Aberrant}) and the single construct standard curve (pCR2.1 HBB cDNA^{Normal} cDNA^{Aberrant}). Orange arrowed box indicates the orientation of the HBB cDNA^{Normal} fragment, whereas the orange arrowed box with the yellow stripe indicates the orientation of the HBB cDNA^{Aberrant} fragment with yellow to correspond to the 19nt misspliced sequence.

Accurate quantification of the two variants is achieved with the use of plasmid standard curves. Initially, two separate constructs were developed, each holding the cDNA fragment from the correctly or aberrantly spliced HBB mRNA (Figure 20) (see Materials and Methods, section 2.6.8). A standard curve from serial dilutions of both plasmids together in a 1:1 ratio was prepared. Although the sensitivity and efficiency of the reaction was good, the peaks for the two variant-specific constructs did not fall together as we expected (Figure 21). The use of separate constructs for the preparation of a single standard curve is permissive to pipetting errors during the preparation of the SC dilutions. Therefore, to avoid these errors, both cDNA fragments were finally cloned in the same construct (see Materials and Methods, section 2.13).

A comparison of the two standard curves (SC), double-plasmid versus single-plasmid SC, in a multiplex qPCR reaction, showed the superiority of using a single-plasmid SC, in terms of accuracy (similar Ct values for both cDNA variants), sensitivity (detection down to ~40 molecules) and reproducibility (Figure 21). Equations from the single-construct SC were extracted based on the

obtained Ct values and known quantities of normal and aberrant curves (Figure 22). External standard curve equations can be used for the re-quantification of samples run on experiments with the standard curves with two-plasmid SC. Nevertheless, this is applicable only in cases where the samples and respective controls were run on the same reaction plate.

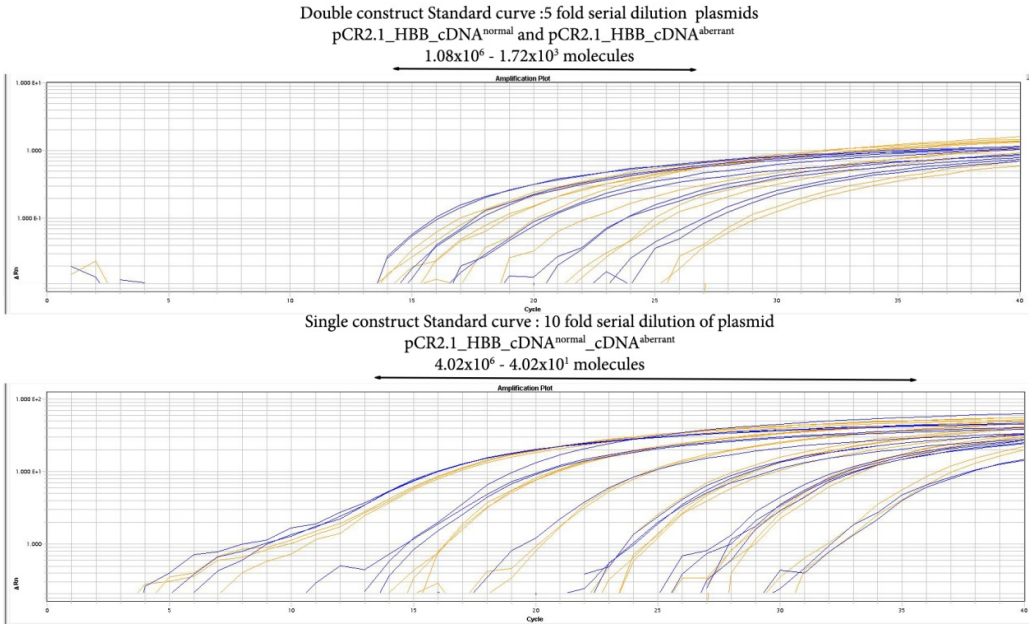


Figure 21: Comparison of amplification plots of double- versus single- plasmid standard curve.
Top: RT-qPCR amplification profiles of the double-plasmid standard curves in which pCR2.1 HBB cDNA^{Normal} and cDNA^{Aberrant} constructs were used in a 1:1 ratio in a range of known quantities ($1.08x10^6$, $2.15\ x10^5$, $4.31\ x10^4$, $8.61\ x10^3$ and $1.72\ x10^3$ molecules/reaction). Bottom: Amplification profiles of single-plasmid standard curves in which pCR2.1 HBB cDNA^{Normal} – cDNA^{Aberrant} construct was used in a range of known quantities ($4.02\ x10^6$, $4.02\ x10^5$, $4.02\ x10^4$, $4.02\ x10^3$, $4.02\ x10^2$ and $4.02\ x10^1$ molecules/reaction). The blue and orange curves represent the detection of amplification of the normal and aberrant cDNA construct, respectively.

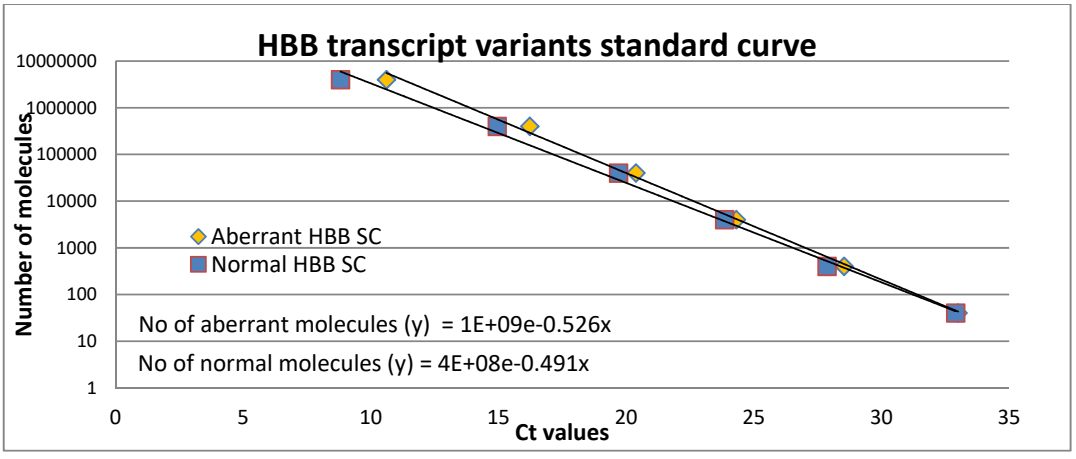


Figure 22: Single-construct SCs and external equations for HBB variant quantification.
In this graphical representation of single-construct standard curves of known quantities (y axis) with the corresponding detected Ct values (x axis) after multiplex RT-qPCR, extracted equations for both HBB variants are shown in orange and blue for the aberrant and normal HBB variants, respectively. External equations were used only for the re-quantification of samples that were analysed with the double construct standard curve.

3.1.2.2 *Generation and characterisation of murine erythroleukaemia transgenic control cells lines*

In general the final cell targets of both the personalised GT approaches to be developed are patient-derived human haematopoietic CD34^{+} progenitor cells. However, before proceeding to application of our tools to CD34^{+} cells, testing of the methods on erythroid cells lines carrying *HBB*$^{IVSI\text{-}110\ (G>A)}$ was deemed necessary due to limited availability of these primary human cells. Besides the humanised *HBB*$^{IVSI\text{-}110}$ mouse model system [90], there are no commercially available cell lines expressing *HBB*$^{IVSI\text{-}110\ (G>A)}$. Therefore, work was initiated to generate *HBB*$^{IVSI\text{-}110\ (G>A)}$ and *HBB*^{Normal} expressing model cell lines using the well-studied and characterised murine erythroleukaemia (MEL) cells. MEL cells were chosen because of their ease of handling and ability to be induced to undergo terminal erythroid differentiation [240]. Furthermore, as MEL cells have a murine genetic background transcripts and protein from the human transgenes would be distinguishable from the equivalent endogenously expressed murine *Hbb-b1* and *Hbb-b2*. As described in the Materials and Methods (section 2.6.9), the MA821 *HBB*^{Normal} LV vector (GLOBE), was modified by site directed mutagenesis to MA821 *HBB*$^{IVSI\text{-}110\ (G>A)}$ (Figure 23) with the mutant construct confirmed by Sanger sequencing (Figure 24). MEL cells were transduced with the MA821 HBB^{Normal} and MA821 HBB$^{IVSI\text{-}110\ (G>A)}$ LVs with a range of MOIs. Transduced cell pools were then selected with an average VCN ~2 for both constructs by employing a multiplex qPCR assay, which was painstakingly established by Ioanna Christodoulou with the aid of the MEL clones and a selection of genomic DNAs described here and produced by me [235]. The MA821 HBB$^{IVSI\text{-}110}$ VCN ~1.9 MEL pool was used as model system for the assessment of functional correction by genome editing in bulk cell experiments, whereas the MA821 HBB^{Normal} VCN~2 cell pool served the "normal" +ve control. In addition, MEL cell clones harbouring MA821 HBB$^{IVSI\text{-}110}$ were produced with VCN 1, 2, 3 and 4 after two rounds of clonal selection. VCNs were quantified by qPCR [235]. An MEL MA821 HBB$^{IVSI\text{-}110}$ clone with VCN 1 was used as model cell line for functional analysis of the gene addition approach (section 3.2.5) and functional correction by genome editing at the clonal level (section 3.3.6).

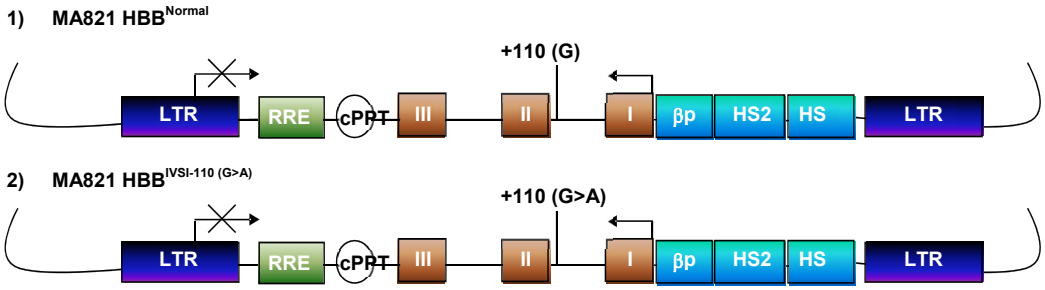


Figure 23: Provirus schematic of MA821 HBB^{Normal} (1) and HBB$^{IVSI-110}$(G>A) mutated version (2).
LCR of HBB; βp, β-globin promoter; HS2 and HS3, hypersensitive site elements 2 and 3 of the LCR; LTR, long terminal repeat with self-inactivating (SIN) design (Deletion of the U3 region of the 3' LTR) ; RRE, Rev response element; cPPT, central polypurine tract; I, II, III, exons 1 through 3; +110 (G>A), IVSI-110 (G>A) misspliced mutation inserted to the MA821 HBB^{Normal} construct via *in vitro* site-directed mutagenesis.

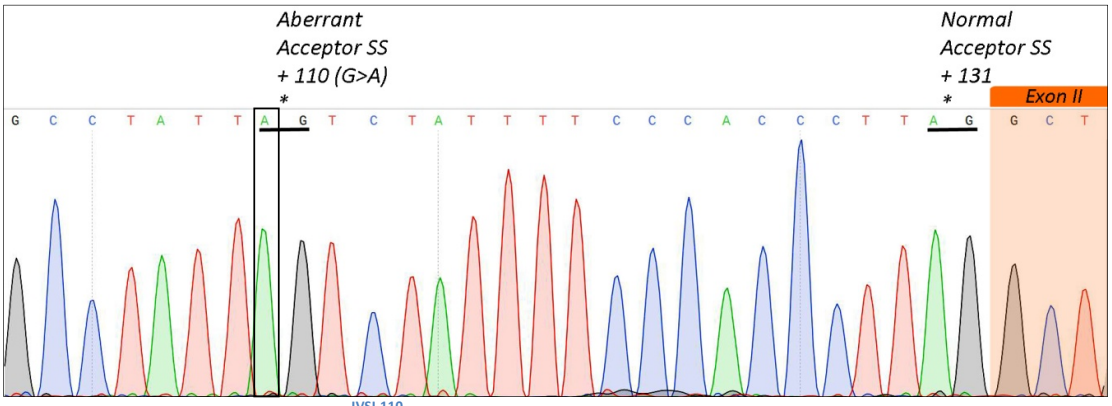


Figure 24: Confirmation of MA821 HBB$^{IVSI-110}$(G>A) LV vector by Sanger sequencing.
Confirmation of bacteria plasmid clone with MA821 HBB$^{IVSI-110}$(G>A) construct after *in vitro* site-directed mutagenesis, using as template MA821 HBB^{Normal}. Sequencing trace of the HBB Intron 1 / Exon 2 region (orange box) and in rectangle designates the substitution +110 G>A mutation. Aberrant and normal acceptor splice sites (SS) are designated as underlined AG di-nucleotides.

Using multiplex RT-qPCR (section 3.1.2.1), an accurate measure of the aberrantly and correctly spliced mRNA was obtained in the total HBB mRNA populations in cell cultures of the transgenic MEL cell lines expressing the HBB$^{IVSI-110\ (G>A)}$ and in differentiated patient-derived haematopoietic CD34^{+} progenitor cells. In particular, the transgenic MEL MA821_HBB$^{IVSI-110}$ cell pool with an average VCN ~1.9 (n=7) and the clonal transgenic MEL MA821_HBB$^{IVSI-110}$ cell line with VCN 1 (n=4) showed that approximately ~60% of the mature HBB mRNA is aberrant. Similar percentages were also observed in 7-day differentiation liquid cultures of patients-derived CD34^{+} cells (n=8; three patients) with the fraction of aberrant transcripts constituting ~54.5 % (Figure 25).

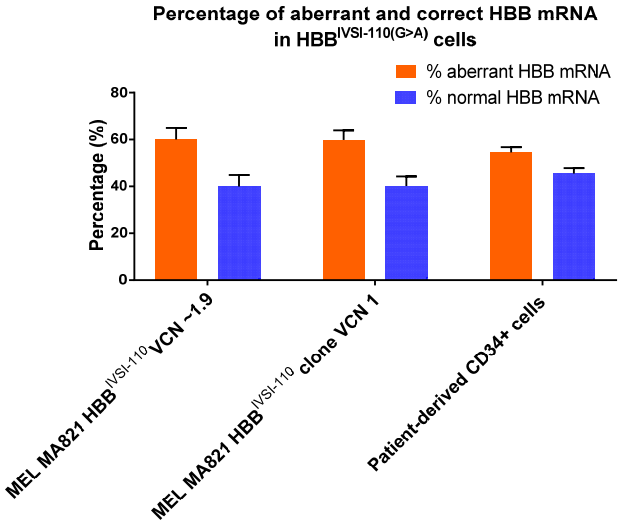


Figure 25: Percentage of correct and aberrantly spliced mRNA in HBB$^{IVSI-110(G>A)}$ transfected cell pools and cell lines.
Percentage of correctly (blue bars) and aberrantly (orange bars) spliced HBB mRNAs in transgenic MEL HBB$^{IVSI-110}$ cell line induced to undergo erythroid differentiation for 3-days with average VCN 1.9 (n=7) and VCN 1 (n=4) and patient-derived CD34^{+} cultures (n=8) on day 7 of differentiation. Note: quantification of the MEL MA821$^{IVSI-110}$ VCN 1.9 was performed with external equation SC, as the particular reactions were performed with the non-accurate double-construct SC. Both humanised MEL HBB$^{IVSI-110}$ cell lines express ~60% of aberrant HBB mRNA, which is similar to the percentage observed in 7-day differentiation cultures of patient-derived CD34^{+} cells in which the percentage of aberrant HBB mRNA is ~54.5%.

The *HBB*$^{IVSI-110}$ gene expression profile at a protein level was measured relative to the MA821 HBB^{Normal}, +ve control. The results from three independent differentiation experiments indicate that both HBB$^{IVSI-110}$ cell lines express approximately 94% less HBB chains compared to the normal +ve control (Figure 26). This findings are in agreement with the expression profile observed in the HBB$^{IVSI-110}$ humanised mouse model system, and most importantly with the amount seen in thalassaemic HBB$^{IVSI-110}$ patients relative to healthy individuals [90]. Therefore, the development of a reliable model cell line, which recapitulates at least at the level of translation HBB expression in PB of HBB$^{IVSI-110}$ thalassaemics was successfully achieved.

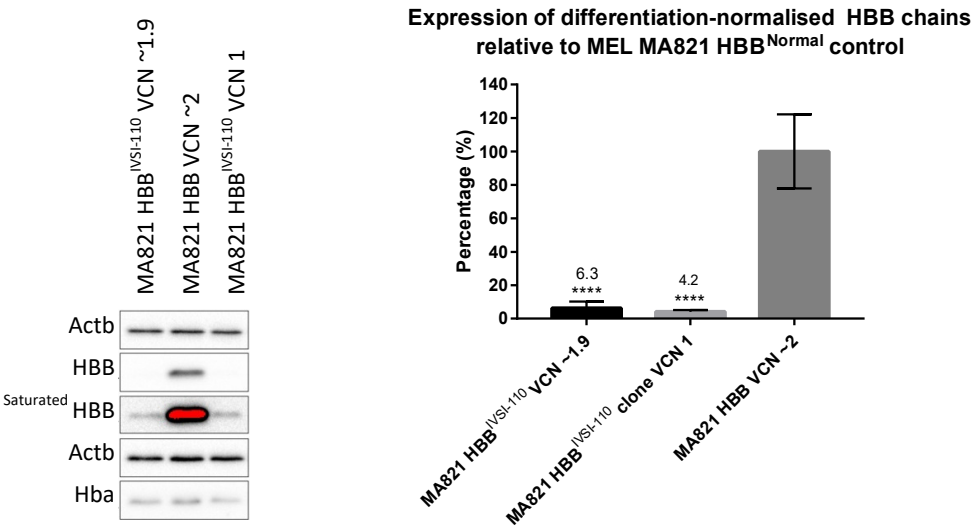


Figure 26: Characterisation of humanized transgenic MEL cell lines at the level of HBB chain production.
Left: Detection of vector-derived human HBB chains, endogenous murine Hba chains and murine Actb on immunoblots of protein extracts from humanised MEL cell lines; MEL MA821 HBB$^{IVSI-110(G>A)}$ average VCN 1.9 and clonal cell line VCN 1 and MEL MA821 HBB^{Normal} average VCN 2. Right: Graphical representation of the percentage of differentiation-normalized HBB chains (HBB chains normalised on Hba chains; HBB/Hba band density ratio) relative to the expression observed in "normal" transgenic cell line MEL MA821 HBB^{Normal} average VCN ~2. Both cells lines express ~94% less HBB chains relative to the cell line with the normal *HBB*. The Actb signal was used as a sample loading control for normalisation purposes based on band quantification of the immunoblot results. Statistical analysis was performed by one-way ANOVA.
Statistical analysis
MEL MA821 HBB^{Normal} VCN ~2 compared to:
MEL MA821 HBB$^{IVSI-110}$ VCN ~1.9 **** *p* value 0.0001
MEL MA821 HBB$^{IVSI-110}$ VCN 1 **** *p* value 0.0001

3.1.3 *Conclusions: Development of model cell lines and quantification methods.*

This section describes the development and characterisation of quantification methods and model cell lines, which are required to address the development of personalised GT for patients harbouring the HBB$^{IVSI-110\ (G>A)}$ mutation, either by gene addition or genome editing.

The data presented indicates the sensitivity and accuracy of the multiplex RT-qPCR for proportional quantification of the normal and aberrant splice mutant HBB mRNA variants. Another advantage of the method developed is that it does not require normalisation for the amount of cDNA against an

endogenous housekeeping gene (e.g., ACTB) or on the level of erythroid differentiation (e.g., HBA induction) since measurement of both normal and aberrant HBB mRNAs takes place in the same reaction and the outcome of the quantification is the relative quantities of the two HBB variants. However, absolute quantities of normal and aberrant splice variants can be obtained by normalising expression levels of total HBB transcripts over Hba expression relative to untreated samples using the ΔΔCt method. This type of quantification was applied for the assessment of restoration of splicing in genome edited MEL MA821$^{IVSI-110}$ transgenic cells at a clonal level (section 3.3.6.1.1). In addition, we have demonstrated that both normal and aberrant variants are measurable, and most importantly the least stable aberrantly spiced mRNA constitutes a significant percentage of the total HBB mRNA pool, even in the case of liquid cultures of CD34^{+} cells derived from thalassaemic patients homozygous for the IVSI-110 mutation (Figure 25).

Furthermore, we have developed murine MEL cells, which contain and express either the *HBB*$^{IVSI-110\ (G>A)}$ or *HBB*^{Normal} transgenes, after transduction with the appropriate version of the MA821 HBB LV, and selected cell populations, which have similar average VCN (~1.9 and ~2 respectively) and clones with known VCN 1, 2, 3 and 4 harbouring the MA821 HBB$^{IVSI-110(G>A)}$.

Transgene expression in the MEL MA821 HBB$^{IVSI-110\ (G>A)}$ transduced cell pool with average VCN ~1.9, MEL clone VCN 1, and MEL MA821 HBB^{Normal} (average VCN ~2) lines were characterised at the mRNA and protein level. We have shown that both cell lines with the HBB$^{IVSI-110(G>A)}$ exhibit close to a 1:1 ratio of the aberrantly (~60%) and normally (~40%) spliced HBB mRNA variants. The similar pattern was also seen in day-7 differentiated patient-derived CD34^{+} cell liquid cultures. In addition, a comparison of the amount of HBB chains levels between the HBB$^{IVSI-110(G>A)}$ and the normal counterpart HBB^{Normal} cell lines following normalisation for differentiation, showed a reduction of ~95% in the mutant cell lines compared to the normal counterpart, which is similar to the reduction of HBB chains seen in IVSI-110 β-thalassaemia patients compared to healthy individuals [90].

In summary, a sensitive and accurate method for the quantification for normal and aberrant spliced HBB mRNA variants was successfully generated, which previously depended on semi-quantitative methods such band density quantification from agarose or polyacrylamide gel electrophoresis. In addition, humanised MEL HBB$^{IVSI-110}$ and HBB^{Normal} transgenic MEL cell lines were successfully generated, which exhibit the protein expression pattern seen in IVSI-110 β-thalassaemia patients.

3.2 Gene addition approach

3.2.1 *Aims*

- Design RNAs specific for the aberrant HBB mRNA
- Assessment the translational advantage of correctly spliced HBB mRNA provided by targeting the aberrant HBB mRNA with shRNA on:
    - Humanised transgenic MEL MA821$^{IVSI\text{-}110(G>A)}$ cell lines
    - Patient-derived hCD34^{+} cells

3.2.2 *Experimental strategy*

As described in the introduction, the common HBB$^{IVSI\text{-}110\ (G>A)}$ missplice mutation leads to the formation of an aberrant cryptic SA site, 19 nt upstream from the normal SA site in intron 1, leading to the incorporation of that fragment to the mature mRNA, which also creates a premature stop codon (PTC; Figure 20). Nevertheless, a small fraction 10–20% of the pre-mRNA is correctly spliced, but the levels of functional HBB chains are not sufficient to be of therapeutic value [82,241]. Although the presence of the PTC on the aberrantly spliced HBB mRNA makes it a target for destruction through the efficient nonsense mediated decay pathway, it has been shown that a relatively large fraction of the HBB mRNA population is aberrantly spliced (close to 1:1 ratio) in liquid cultures of differentiated patient derived CD34^{+} cells (Figure 25).

Hypothesis

We believe that there is a competition for translation factors between the correctly and aberrantly spliced HBB mRNA variants, which may impair the therapeutic potential of *HBB*-containing LVs.

In order to test this hypothesis shRNAs were designed to target the aberrantly spliced HBB$^{IVSI\text{-}110}$ transcript (Figure 27). Since the difference between correct and aberrant mRNA is only a 19 nt misspliced sequence, the shRNA-targeting region limited the design options to three potential candidates with overlapped target sequences. The shIVSI-110 Upstream (Up) and Downstream (Down) partially target a sequence that is also found in correctly spliced mRNA, whereas shIVSI-110 Middle (Mid) is mainly targeting the 19 nt sequence present only in the aberrant mRNA (Figure 27).

By analogy to previous studies targeting SCD mutation on HBB mRNA [158], the key strategy presented here is for the co-expression of vector-derived HBB and three different IVS1–110-specific shRNAs inserted within intron-2 of the vector *HBB* (Figure 29). To this end the junction point of the internal intron-2 deletion of modified anti-sickling MA821-T87Q LV, which is close to the unspecified position previously described Samakoglu *et al.,* was selected [158]. The goal of this design was expression of the

anti-sickling *HBB*^{T87Q} transgene under the regulation of erythroid specific promoter and βLCR, and simultaneous delivery of the shIVSI-110 RNAs sequence within intron 2, leading to knock-down of the aberrant mRNA.

In addition to this design, the specific shRNA-encoding sequences were also cloned independently in the commercially available pLKO.I U6 LV for single or double transduction in combination with MA821 LV (GLOBE LV). U6 is an RNA pol3-specific promoter, and clusters of dT_4 and dT_5 are recognised in mammals as termination signals [242,243]. In particular, the shIVSI-110 Mid-encoding sequence holds a repeat of TTTT, the corresponding U_4 sequence being part of the aberrantly spliced HBB mRNA target sequence, and might impair U6-driven shIVSI-110 Mid expression. Therefore, an additional modified version of shIVSI-110 Mid was designed, shIVSI-110 Mid2, in which the TTTT repeat was substituted with TTGT, whereas the antisense sequence was left unchanged (AAAA), giving an imperfectly paired shRNA stem sequence. The antisense sequence is the part of the shRNA that will be used by DICER as guide strand for the recognition and destruction of the aberrantly spliced HBB mRNA (Figure 28)[244].

All LVs (single pLKO.I- and hybrids MA821 T87Q- shIVSI-110 encoding RNAs) were produced and titrated. However, in parallel work conducted by a colleague in the laboratory (data not shown), who used the same cloning strategy and delivery for knock-down of BCL11A, no silencing was observed even though expression of the *HBB* transgene was not noticeable affected. This implied that the shRNAs embedded within Intron-2 were not being appropriately processed correctly after splicing of the HBB^{T87Q} pre-mRNA. Therefore, before proceeding to a more complex system of expressing simultaneously *HBB* and shRNAs from the same transgene, we decided to perform single transductions with pLKO.I shIVSI-110 RNAs and/or in combination with MA821 HBB^{Normal} LV to establish first a proof of principle of our hypothesis.

Efficacy and safety of the pLKO.I shIVSI-110 vectors was initially tested in clonal transgenic MEL cell lines containing the MA821_*HBB*$^{IVSI-110}$ transgene at VCN 1, and then alone or in combination with MA821 HBB^{Normal} LV in hCD34^{+} HPC cells isolated from the blood of HBB$^{IVSI-110}$-thalassaemic patients and healthy individuals.

Transduced transgenic MEL cells were induced to undergo erythroid differentiation and evaluation of the knock-down effect was assessed at days 3 and 6 post-differentiation at the RNA level by multiplex RT-qPCR, whereas the translational advantage offered by the shIVSI-110 RNAs at a protein level, was quantified on day 6 of differentiation via immunoblots. Cultures of transduced hCD34^{+} cells were differentiated down the erythroid lineage and evaluated for HBB chain production. In both experiments, MEL and hCD34^{+}, VCN was measured by multiplex qPCR.

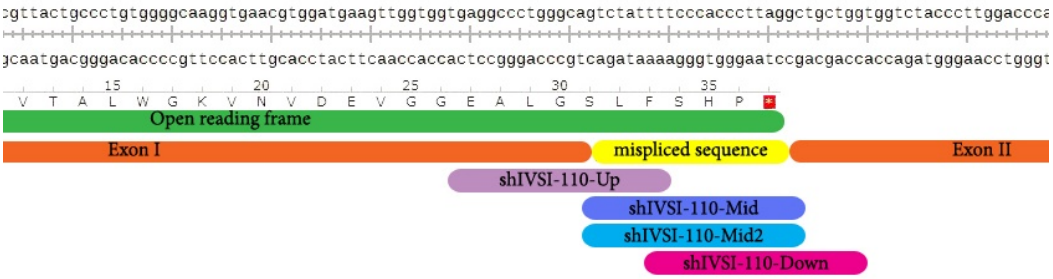
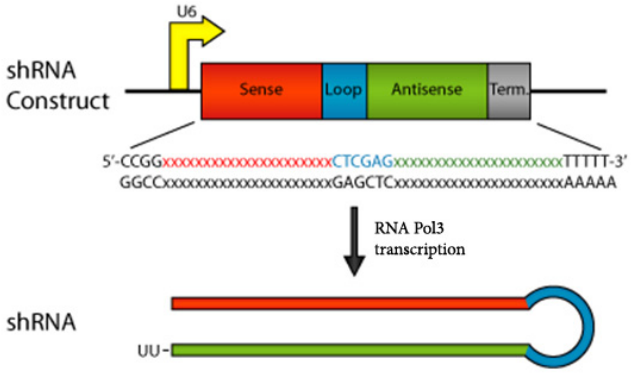


Figure 27: Target sequences of shIVSI-110 RNAs on the aberrantly spliced HBB$^{IVSI-110}$ transcript sequence.
The shIVSI-110 specific RNAs are designated as purple -Up, blue -Mid, light blue Modified -Mid2 and Red -Down. Green box: the open reading frame of abnormal HBB mRNA with corresponding amino acid sequence (with pre-termination codon denoted by the red asterisk). Orange rounded boxes: Exon I/II encoded region, yellow rounded box: abnormally spliced-in sequence (19 nt) from Intron 1 in aberrant mRNA.



shIVSI-110 encoding RNAs	Sense strand	Loop	Antisense strand	Term. signal
shIVSI-110 Up:	ccgg GAGGCCCTGGGCAGTCTATTT	gttaac	AAATAGACTGCCCAGGGCCTC	TTTTTG
shIVSI-110 Mid:	ccgg GTCTATTTTCCCACCCTTAGG	agtact	CCTAAGGGTGGGAAAATAGAC	TTTTTG
shIVSI-110 Mid2:	ccgg GTCTATTGTCCCACCCTTAGG	agtact	CCTAAGGGTGGGAAAATAGAC	TTTTTG
shIVSI-110 Down:	ccgg TTTCCCACCCTTAGGCTGCTG	agtact	CAGCAGCCTAAGGGTGGGAAA	TTTTTG

Figure 28: shIVSI-110 RNA sequences.
Top: pLKO.I shIVSI-110 RNAs LV encoding cassette and the produced shRNA hairpin like structure. shRNAs are driven under RNA pol3 U6 promoter. shRNA sequence comprised by the sense strand (red), the antisense (green) which are linked with a 6bp palindromic sequence (blue) usually constituting a restriction enzyme site, used for the confirmation of the positive constructs with restriction enzyme digest or to allow the full sequencing of the construct that might be block due to the secondary structures at the site of the loop. Bottom: shIVSI-110-specific RNA sequences (Up, Mid, Mid2 and Down). Rectangle designates the substitution of TTTT to TTGT in the modified version of Mid, Mid2, to avoid possible recognition of the TTTT repeat as termination signal (Term. Signal), TTTTT, by RNA pol3.

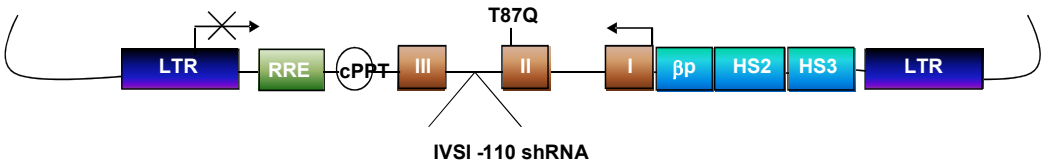


Figure 29: Schematic of MA821-T87Q and shRNA cloning site within intron-2.
MA821-T87Q LV is the modified version of the MA821 HBB (GLOBE) which holds the anti-sickling T87Q mutation. LTR of HBB; T87Q, threonine-to-glutamine codon change; βp, *HBB* promoter; HS2 and HS3, hypersensitive site elements 2 and 3 of the βLCR; LTR, long terminal repeat with self-inactivating (SIN) design (Deletion of the U3 region of the 3' LTR); RRE, Rev response element; cPPT, central polypurine tract; I, II, III, exons 1 through 3 of *HBB*.

3.2.3 *Development of shRNA-encoding LV constructs*

3.2.3.1 *Overview of the cloning strategy*

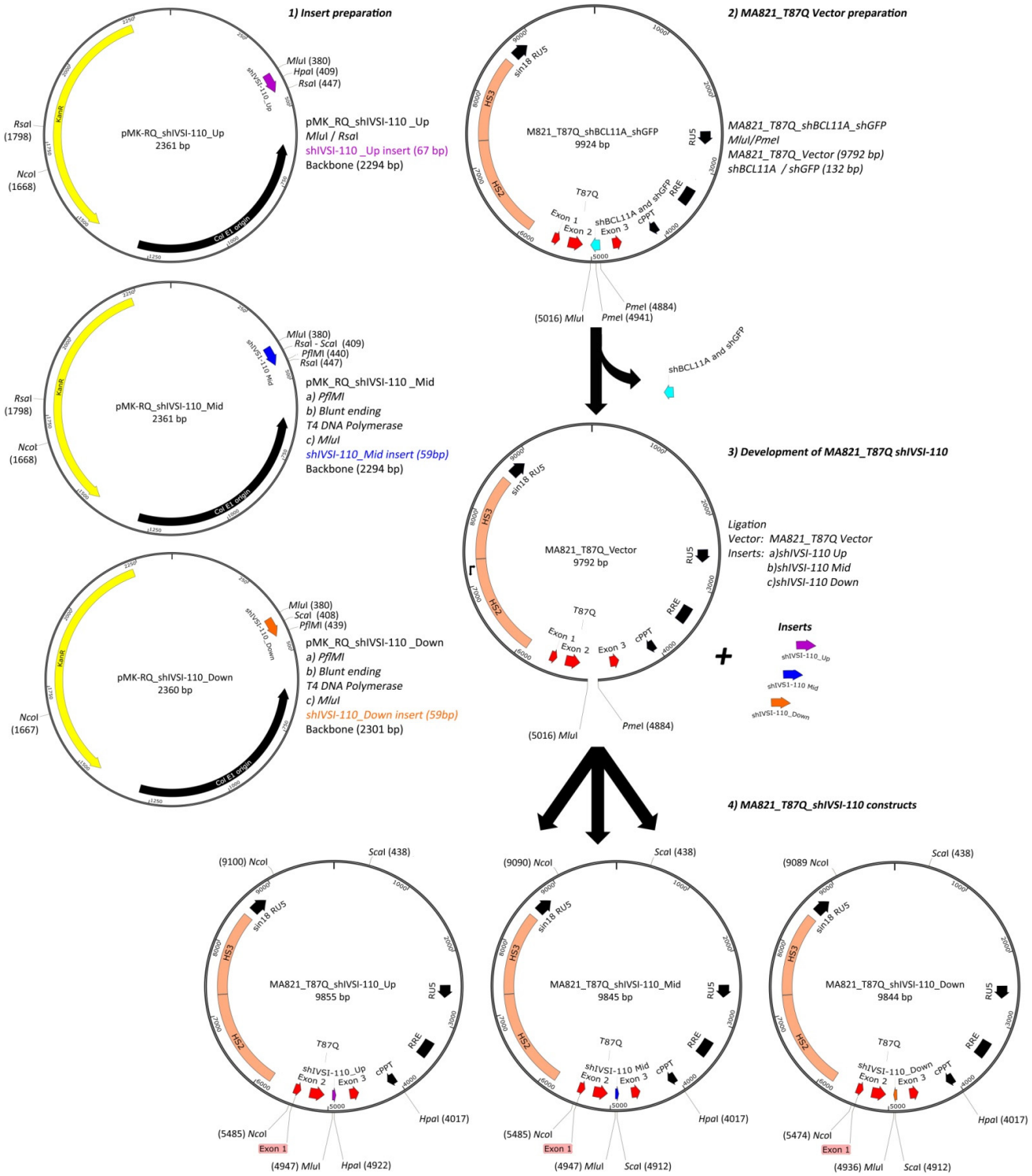


Figure 30: Cloning strategy for MA821-T87Q-shIVS1-110 LV constructs.
1) Insert preparation: shIVSI-110 RNAs (Up, Mid and Down) were received in pMK_RQ_plasmid contructs. The shIVSI-110_Up insert (purple arrow) was released from the pMK_RQ_shIVS1-110 plasmid with *Mlu*I/*Rsa*I, creating compatible ends for cloning into MA821-T87Q vector. The shIVSI-110 Mid (blue arrow) and Down (orange arrow) inserts were produced by *PflM*I digestion, blunted with T4 DNA Pol and subsequent digest with *Mlu*I. 2) MA821-T87Q vector was prepared with the removal of the shRNA-encoding DNA (shBCL11A and shGFP, light blue arrow) from MA821-T87Q-shBCL11A_shGFP construct (unpublished data) by *Mlu*I/*Pme*I restriction digestion. 3 and 4) Cloning of shRNA inserts into MA821-T87Q vector produced three novel MA821-T87Q LV constructs, with each holding a unique shRNA sequence specific for aberrantly spliced HBB IVSI-110 mRNA. Kan^R, Kanamycin resistance; Col E1, Origin of replication; T87Q, threonine-to-glutamine codon change; HS2 and HS3, hypersensitive site elements 2 and 3 of the βLCR; sin18 RU5 LTR, long terminal repeat with self-inactivating (SIN) design (Deletion of the U3 region of the 3' LTR); RRE, Rev response element; cPPT, central polypurine tract; I, II, III, exons 1 through 3 of *HBB*.

Using the MA821-T87Q-shBCL11A-shGFP plasmid (data unpublished) as a starting construct, commercially synthesised shRNA-encoding DNA fragments were inserted into the junction site of the vector-encoded *HBB*, as shown in Figure 30.

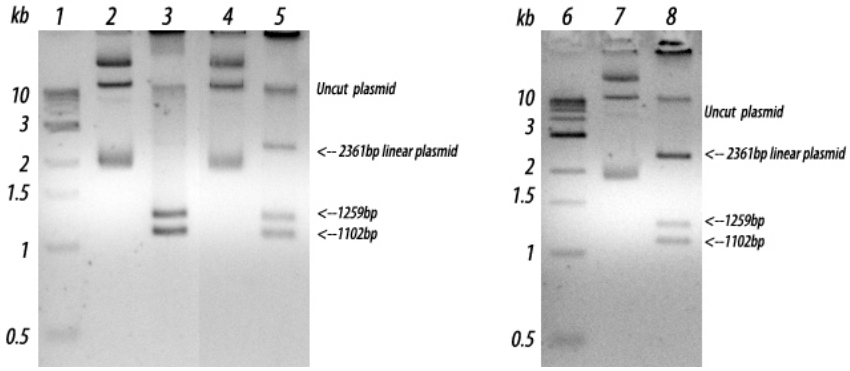


Figure 31: Restriction enzyme digest confirmation of commercial pMK_RQ_shIVS1-110 constructs.
Gel separation of the restriction digestion products of pMK_RQ_shIVSI-10 Up with *Nco*I/*Hpa*I (3) and of pMK_RQ_shIVSI-110 Mid (5) and pMK_RQ_shIVSI-110 Down (8) with *Nco*I/*Sca*I, including undigested control (lanes 2, 4, 7), respectively, and giving the diagnostic bands of 1259 and 1102 bp and the 2361-bp band of the linearised plasmid upon partial digestion. Lanes (1, 6): 1 kb DNA ladder. Maps of the constructs are displayed in Figure 30.

3.2.3.2 *Confirmation of commercial shRNA source clones*

Plasmids encoding IVSI-110-mRNA-specific shRNAs (upstream, middle and downstream) were commercially produced and received separately in pMK_RQ Kan^R vectors flanked by two *Sfi*I restriction sites, as 2361-bp, 2360-bp and 2360-bp constructs, respectively (Figure 30). These constructs were cloned, confirmed by restriction enzyme digestion (Figure 31), expanded to maxi cultures for DNA extraction and stored at -80 °C as bacterial glycerol stocks.

3.2.3.3 *Cloning of shRNA-encoding transfer-vector plasmids*

3.2.3.3.1 MA821_HBB^{T87Q} LV

MA821_T87Q vector (Amp^R) (9792 bp) was prepared by double restriction enzyme digestion of the LV construct MA821_T87Q_shBCL11A_shGFP (Amp^R) (9924 bp) with *Mlu*I and *Pme*I (blunt end). This digestion removed the two shRNAs creating compatible flanking ends to accept the synthetic sequences encoding IVSI-110 shRNAs by ligation using standard procedures. Similarly, the shRNA inserts were prepared accordingly as shown in Figure 30, with an initial *Mlu*I and *Rsa*I (blunt end) digestion to release the shIVSI-110_Up insert and a sequential *PflM*I digest, blunt-ending and *Mlu*I digest for the shIVSI-110_Mid and shIVSI-110_Down inserts. The bands of interest were isolated and ligated before bacterial transformation and confirmation of individual clones by plasmid mini-prep cultures and analytical restriction digests as shown for MA821-T87Q-shVSI-110-Up as a representative example in Figure 32 (other data not shown).

3.2.3.3.2 pLKO.I shRNA LV

Cloning of shIVSI-110-encoding RNA in pLKO.I vector was performed by insertion downstream of the U6 promoter using *EcoR*I and *Age*I cloning sites. Positive clones were selected following appropriate restriction enzyme digests targeting the unique site on the loop of each shIVSI-110 insert, *Hpa*I for shIVSI-110-Up and *Sca*I for Mid/Mid2 and Down, leading to linearization of the constructs (Figure 33). Positive clones were confirmed by Sanger sequencing (data not shown).

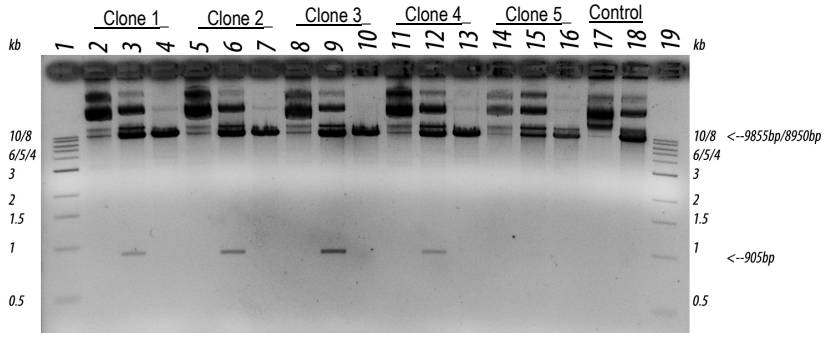


Figure 32: Confirmation of MA821-T87Q-shIVS1-110-Up plasmid clones.
Candidate clones for MA821-T87Q-shIVS1-110 were digested with uncut, *Sca*I-digested and *Hpa*I-digested DNA shown for each clone, giving, respectively, a retained supercoiled plasmid, a diagnostic 905-bp fragment and a linearised 9855-bp fragment for each correct clone. Lanes (17, 18) – MA821-T87Q-B1G control, uncut and *Hpa*I-digested, respectively.
Lanes (1, 19) – 1-kb DNA ladder. Maps of the constructs are displayed in Figure 30.

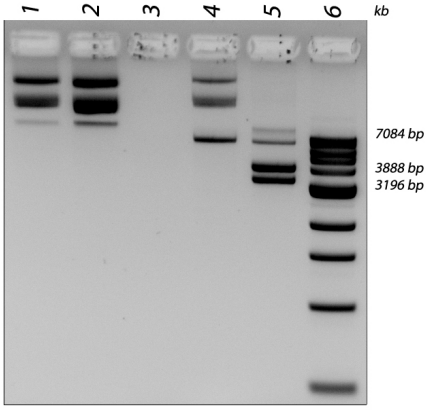


Figure 33: Confirmation of pLKO.I-shIVS1-110-Up/MID2 plasmid clones.
Candidate clones for pLKO.I-shIVS1-110- Up and Mid were confirmed by digestion with **Sca**I and **Hpa**I respectively, with uncut (1,2), **Hpa**I-digested (4) and **Sca**I-digested (5) DNA giving, respectively, a retained supercoiled plasmid, a linearised 7084-bp fragment and two fragments 3888/3196bp for each correct clone. Lanes (1, 2): uncut pLKO.I shIVSI-110 Up and Mid2, respectively. Lanes (4, 5): HpaI and ScaI digested DNA of pLKO.I shIVSI-110 Up and Mid2, respectively
Lanes (6): 1-kb DNA ladder

3.2.4 *Titration of shRNA-encoding LVs in HEL cells*

After confirmation and large-scale plasmid production of all the required transfer-vector plasmids, LV stocks were produced in HEK 293T cells and biological viral titres determined by transduction of HEL cells (section 2.20.3). The biological titres were assessed by qPCR at day-20 post-transduction. Standard curves, primers and probes were as described (see section 2.20.3.1). All titres were calculated according to the formula given in 2.20.3.1 and results are shown in Table 9. In general,

high titres were obtained for the entire set of LVs with an average of $5.3x10^8$ transducing units/mL (TU/mL). Concentration and titration of the MA821 $HBB^{IVSI-110\ (G>A)}$ LV was not required since the specific vector was used only for the production of a low VCN cell line described in section 2.22. Therefore, transduction with serial dilution of the un-concentrated LV was sufficient to establish a cell line with average VCN ~1.9.

Table 9: Titres of lentiviral vectors.
Results of titration of LVs by multiplex qPCR. Titres are shown as transducing units/mL of each LV. MA821 $HBB^{IVSI-110(G>A)}$ LV was not titrated, since it was used specifically for the production of the MEL MA821 $HBB^{IVSI-110}$ average VCN ~1.9 cell line. Dilutions of the unconcentated LV stock were used to obtain low average VCNs for the production of the MEL MA821 $HBB^{IVSI-110}$ cell line and thus the concentration of the LV stock was not required.

LV	qPCR titre (TU/mL)
MA821 HBB^{Normal}	$8.79x10^7$
MA821_HBB^{T87Q}	$1.69x10^8$
MA_821_ HBB^{T87Q} _shIVSI-110_Up	$2.40x10^8$
MA_821_ HBB^{T87Q} _shIVSI-110_Mid	$4.31x10^8$
MA_821_ HBB^{T87Q} _shIVSI-110_Down	$2.71x10^8$
pLKO.I shIVSI-110 Up	$5.77x10^8$
pLKO.I shIVSI-110 Mid	$3.92x10^8$
pLKO.I shIVSI-110 Mid2	$2.30x10^8$
pLKO.I shIVSI-110 Scramble	$2.30x10^9$
pLKO.I shIVSI-110 Down	$7.76x10^8$
MA821 $HBB^{IVSI-110\ (G>A)}$	N/A

3.2.5 *Characterisation of LVs in transgenic MEL MA821_$HBB^{IVSI-110}$ cells*

pLKO.I shIVSI-110-encoding LVs (Up/Mid/Mid2/Down) were first tested on the clonal transgenic MEL MA821_$HBB^{IVSI-110}$ (VCN 1) cell line in three separate experiments, using as negative controls transduction with pLKO.I shScramble and untransduced Mock control. A pilot experiment (Exp1) was performed where a range of MOIs (5, 10, 20 and 40) were used for the pLKO.I shRNA LVs in order to evaluated the appropriate MOI that will give similar VCNs (data not shown). Cells were left to recover for 14-20 days before collecting for gDNA and assessment of the VCN via the multiplex qPCR method (section 2.20.3.2). Transduced cells with similar VCN were selected for induced erythroid differentiation and the following two experiments (Exp2 and Exp3) were performed using the same MOIs (Table 10). Worth noting is the extensive cell death observed in the samples transduced with pLKO.I shIVSI-110_Up in all three experiments, even though in the 2nd and 3rd experiment we also performed transductions with lower MOI 5 (shIVSI-110 Up *), in an attempt to reduce the level of toxicity. Although the 2nd and 3rd experiments have similar VCNs, there are 10 to 20 times lower than the VCNs in the 1st pilot experiment (Figure 34). This may be due to transduction in experiments 2 and 3 being performed on different batches of MEL MA821$^{IVSI-110}$ cells, which could have been less permissive to LV transduction. Moreover, apart from the high cell death post-transduction in

samples with pLKO.I shIVSI-110 Up, the recovered population exhibit close to background VCN = 1 (untransduced), which implied that most of the transduced cells with the pLKO.I shIVSI-110 Up died.

Table 10: pLKO.I shIVSI-110 and shScramble multiplicity of infections used in MEL experiments.
The same MOIs were used for the transduction of $10x10^6$ MEL MA821 $HBB^{IVSI-110}$ VCN 1 transgenic cells in experiments. In addition of MOI 10 used for pLKO.I shIVSI-110 Up in all the experiments (Experiments 1–3), we used also MOI 5 in experiments 2 and 3 in an attempt to reduce the observed increased toxicity seen with former MOI.

	MOI		
pLKO.I Lentivirus	5	10	20
pLKO.I shIVS1-110 Up	x	x	
pLKO.I shIVS1-110 Mid	x		
pLKO.I shIVS1-110 Mid2	x		
pLKO.I shIVS1-110 Down		x	
pLKO.I shScramble			x

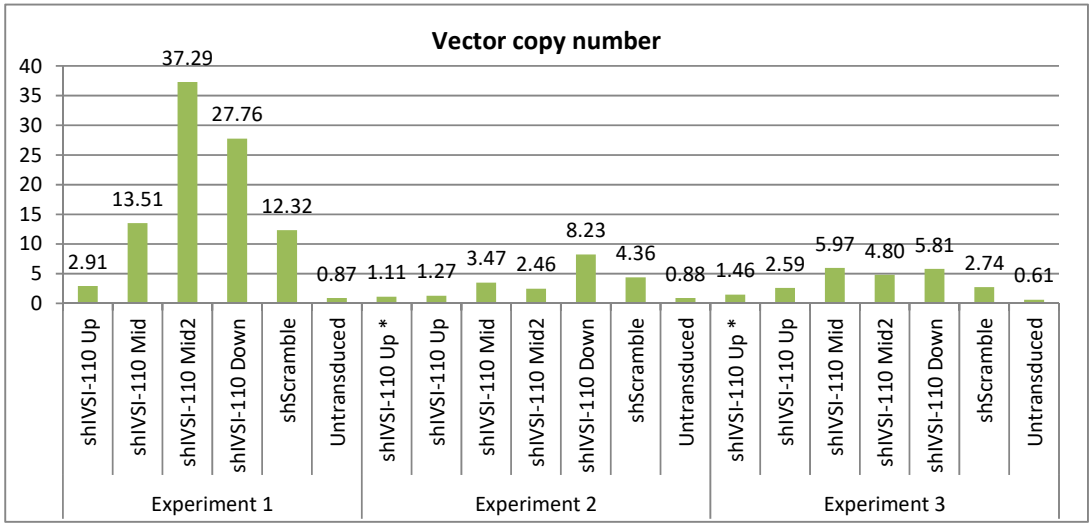


Figure 34: Vector copy number of transduced transgenic MEL MA821 $HBB^{IVSI-110}$ clonal cell line (VCN 1).
VCN of MEL MA821 $HBB^{IVSI-110}$ clonal cell line after transduced with pLKO.I shIVSI-110 (Up, Mid, Mid2 and Down) LVs and pLKO.I shScramble in three independent experiments (exp 1–3). VCN was determined after ~20 days of culture post-transduction by multiplex qPCR. PLKO.I shScramble and untransduced cells (Mock) served the negative controls of the experiments. Cells were transduced with the same MOI in all three experiments, with exception of the pLKO.I shIVSI-110 Up which beside the transduction of cells with MOI 10 (exp 1–3), additional transductions with MOI 5 (designated as shIVSI-110 Up *) were performed in experiment 2 and 3 in an attempt to reduce the observed high cell death.

3.2.5.1 *Functional analysis of pLKO.I shIVSI-110 transduced MEL MA821 $HBB^{IVSI-110}$ cells.*

MEL MA821 $HBB^{IVSI-110}$ cells transduced with LVs encoding pLKO.I shIVSI-110 Up/Mid/Mid2 and Down were differentiated in induction medium (cRPMI supplemented with 1.5% DMSO) along with the pLKO.I shScramble-transduced and untransduced (mock) negative controls, as described in section 2.21. Cells were collected for RNA extraction on day 3 and 6 and for protein on day 6.

3.2.5.1.1 Effect of shIVSI-110 RNAs on transcript levels in transduced MEL MA821$^{IVSI-110}$ clonal cells

The silencing efficiency of the shIVSI-110-specific RNAs against the aberrant HBB mRNA was then assessed using the multiplex RT-qPCR plasmid-based method (section 3.1.2.1).

In general, analysis of mRNA data as a percentage of correctly and aberrantly spliced HBB mRNAs in the total HBB mRNA population on days 3 and 6 following terminal erythroid differentiation, showed that only the pLKO.I shIVSI-110 Down transduced cells in the three experiments exhibited silencing of the aberrant mRNA relative to the negative controls, shScramble and untransduced cells. In particular, cells transduced with pLKO.I_shIVSI-110 Down on day-3 showed a decrease in the average fraction of the aberrant mRNA from 71.3% to 50.3% and on day-6 from 76.4% to 34.7%. The decrease of the fraction of aberrantly spliced HBB mRNA and therefore a corresponding increase in the level of the normal transcript was more pronounced at day-6 of differentiation (Figure 35). Furthermore, statistical analysis of the ratios of correct/aberrant (c/a) HBB mRNA in all three experiments, showed a significant increase in this ratio of the samples transuded with pLKO.I shIVSI-110 Down LV compared to the corresponding untransduced negative controls from 0.4 to 1.0 (p value = 0.0136) and from 0.31 to 2.0 (p-value < 0.0001) at day-3 and day-6, respectively (Figure 36). A similar pattern is evident when data are calculated as the percentage change in the fraction of correctly or aberrantly spliced HBB mRNAs compared to the untransduced cells. These results show a significant decrease of the aberrant mRNA by 29.6% (p value = 0.0003) and by 54.2% (p value = 0.0001) at the day-3 and -6, respectively, in samples transduced with the pLKO.I shIVSI-110 Down LV compared to the untransduced controls (Figure 37).

In summary, we have shown a significant knock-down effect of the aberrant mRNA only in the case of cells transduced with the pLKO.I shIVSI-110 Down LV. However, this analysis only indicated proportional changes of the two variants, not the absolute quantities of the two variants. Therefore, although a silencing effect is observed when the shIVSI-110 Down is used, we cannot discern whether it targets only the aberrant mRNA or whether it targets both mRNA variants, but the aberrant species with higher efficiency.

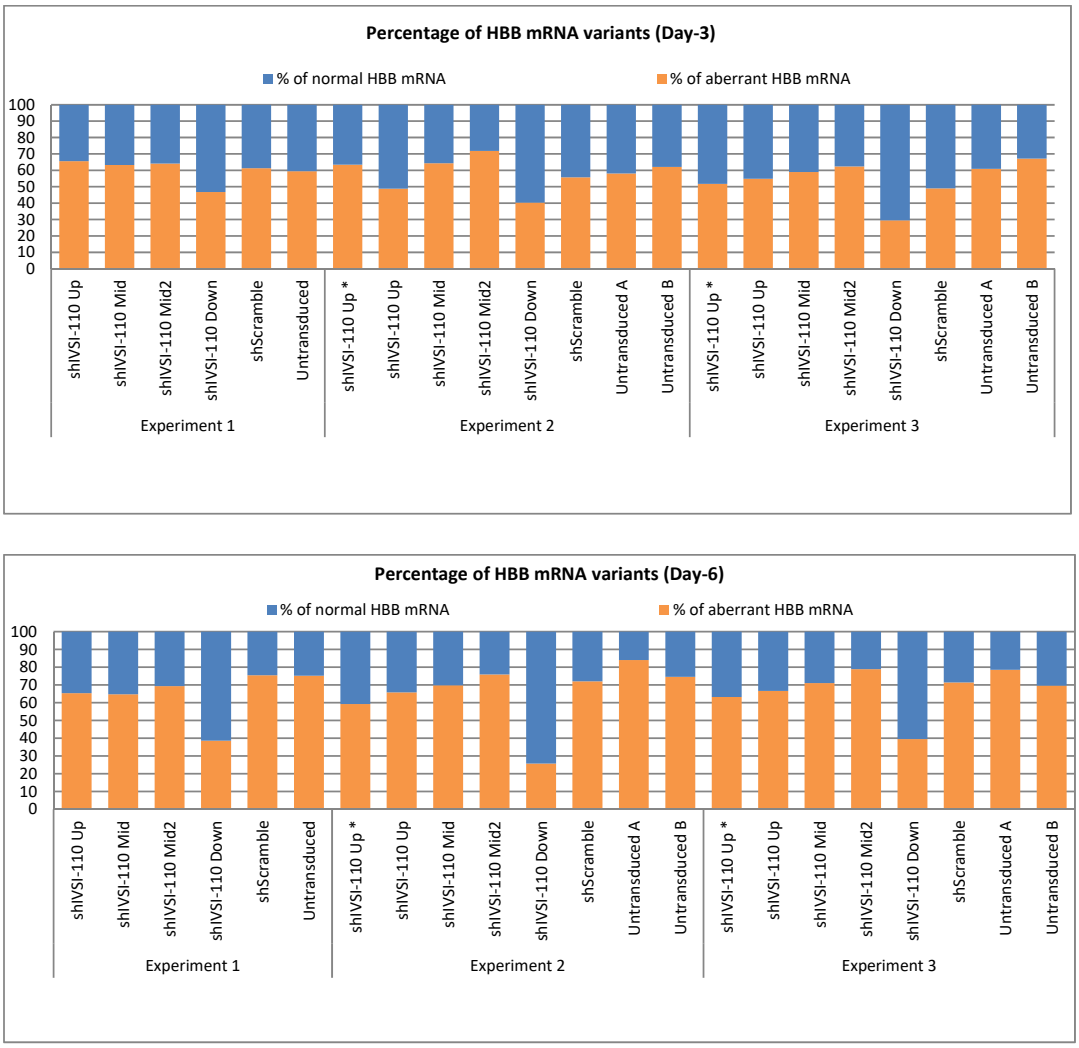


Figure 35: The pLKO.I shIVSI-110 Down LV gives a major decrease in the percentage of HBB mRNA in variants in MEL MA821 HBB$^{IVSI-110}$ cells.

MEL MA821 HBB$^{IVSI-110}$ cells were transduced with the pLKO.I shIVSI-110-encoding LVs (Up/Mid/Mid2/Down), and induced to undergo terminal erythroid differentiation. Ratios and percentages of correctly and aberrantly spliced HBB mRNA were then determined by RT-qPCR at days 3 and 6 post-differentiation. Correctly (blue bars) and aberrantly (orange bars) spliced HBB mRNA levels in each sample at day-3 (top panel) and day-6 (bottom panel) of differentiation in three independent experiments is shown. Transduced samples with pLKO.I shIVSI-110 Up at MOI 10 (designated as shIVSI-110 Up; experiment 1–3) and at MOI 5 (designated as shIVSI-110 Up *; experiment 2–3).

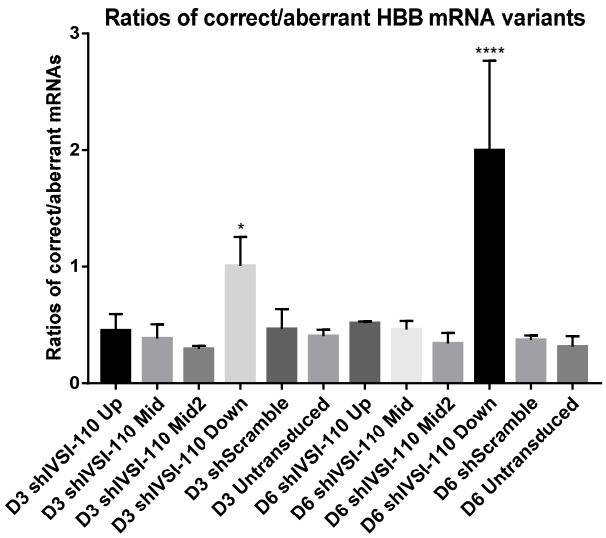


Figure 36: Evaluation of knock-down efficiencies by shIVSI-110 RNAs in MEL HBB$^{IVSI-110}$ VCN1 on day-3 (D3) and day-6 (D6) of erythroid differentiation in three independent experiments (n 3).
mRNA data are shown as the ratios of correct/aberrant HBB mRNA and statistical analysis was performed on transduced samples compared the corresponding day untransduced control. Only pLKO.I shIVSI-110 Down transduced cells exhibited a significant increase of the c/a ratio relative to the control, suggesting the knock-down of the aberrant mRNA. . Statistical analysis was performed by one-way ANOVA.
Statistical analysis:
D3 shIVSI-110 Down * – *p* value 0.0136
D6 shIVSI-110 Down **** – *p* value < 0.0001

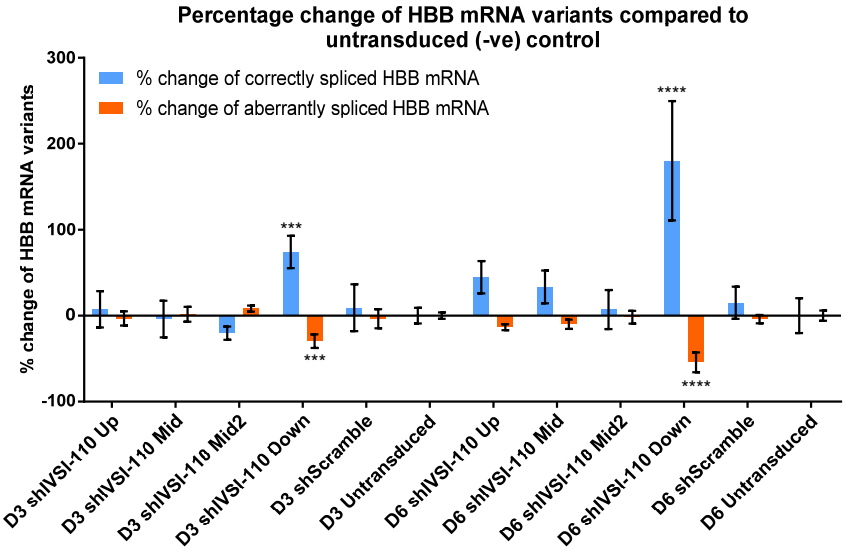


Figure 37: Percentage change of HBB mRNA variants compared to untransduced controls.
Percentage change of correctly (c) (blue bars) and aberrantly (a) (orange bars) spliced HBB mRNAs in pLKO.I shRNA transduced MEL HBB$^{IVSI-110}$ VCN1 cell samples relative to untransduced negative control levels at day-3 (D3) and day-6 (D6) of erythroid differentiation in three experiments (n 3). Statistical analysis was performed on transduced samples compared the corresponding day untransduced control. Statistical analysis was performed by one-way ANOVA.
Statistical analysis:
Untransduced cells compared to:
D3 shIVSI-110 Down *** (c) – *p* value 0.0003
D3 shIVSI-110 Down *** (a) – *p* value 0.0003
D6 shIVSI-110 Down **** (c) – *p* value 0.0001
D6 shIVSI-110 Down **** (a) – *p* value 0.0001
* $p < 0.05$, ** $p < 0.01$*** $p < 0.001$, **** $p < 0.0001$.

3.2.5.1.2 Effect of shIVSI-110 RNAs on HBB protein levels in transduced MEL MA821$^{IVSI-110}$ clonal cells

Differentiated transduced MEL MA821 HBB$^{IVSI-110}$ cells were harvested on day-6 for protein extraction and analysis of HBB chain expression on immunoblots based on band density quantification (section 2.14.3). The levels of HBB protein expression is normalised by reference to the level of degree of differentiation (levels of murine Hba1/2 chains). Also, staining for Actb was used as a protein loading control. All three independent experiments displayed the same banding pattern for HBB, with a clear noticeable increase in samples transduced with the two pLKO.I shIVSI-110 Mid and Mid2 LVs (Figure 38). Quantification and normalisation of band intensities showed an increase in the HBB/Hba ratio of 4-37 fold compared to the untransduced negative controls. In addition, HBB/Hba levels between samples transduced with the pLKO.I shIVSI-110 Mid and Mid2 LVs within the same experiment were similar, indicating that the TTTT > TTGT substitution did not have any profound effects on the expression of shRNA or activity. Samples transduced with the pLKO.I shIVSI-110 Down LV displayed an increased trend of up to 5 fold, whereas the shIVSI-110 Up transduced samples remained closed to background levels (Figure 38). As mentioned previously, the shIVSI-110 Mid and Mid2 are the only shRNAs that specifically target aberrantly spliced HBB mRNA (Figure 28).

In order to analyse protein data from all three independent experiments together, data were first normalized by setting at 100% the sample with the highest value for each tested parameter, HBB/Hba or levels of differentiation Hba, in each experiment. Protein analysis of data from all three experiments together showed, as expected, that samples transduced with the pLKO.I shIVSI-110 Mid and Mid2 present an increase in HBB/Hba levels by an average of 73.56% (p value = 0.0001) and 88.8% (p value = 0.0001) respectively, compared to the levels observed in the untransduced negative control (Figure 39). Samples transduced with the pLKO.I shIVSI-110 Down also showed an increase of 16.1% but not with statistical significance, whereas the shIVSI-110 Up transduced cells remained at the level of the negative control background. Likewise, analysing samples for the assessment of the level of differentiation based on the amount of Hba chains, indicated that among all the samples only the untransduced cells had significantly higher levels of differentiation compared to the shIVSI-110 Mid2 (by 55.1 %; p value = 0.0135), shIVSI-110 Down (by 46.5 %; p value = 0.0423) and shScramble (by 52 %; p value = 0.0205) (Figure 39).

These data indicate that both shIVSI-110 Mid and Mid2 increased significantly the synthesis of functional HBB chains in transgenic MEL MA821 HBB$^{IVSI-110}$ VCN 1 clonal cells even though the ratio of the two HBB mRNA variants remained stable. Importantly and to our surprise, for shIVSI-110 Mid

and Mid2, the observed increase of HBB chains derived solely from the translation of the endogenous normal HBB mRNA and not from the addition of extra functional *HBB* transgene to the cells. These immunoblot data show a discrepancy between the significant increase of HBB transcript ratios for shIVSI-110 Down and a lower increase in protein levels compared to shIVSI-110 Mid and Mid2. The explanation for this phenomenon is unclear at present but might be found in the target site of shIVSI-110 Down, which might affect mRNA processing and prevent abnormal splicing, while also lowering the overall amount of processed mRNA, which was not assessed.

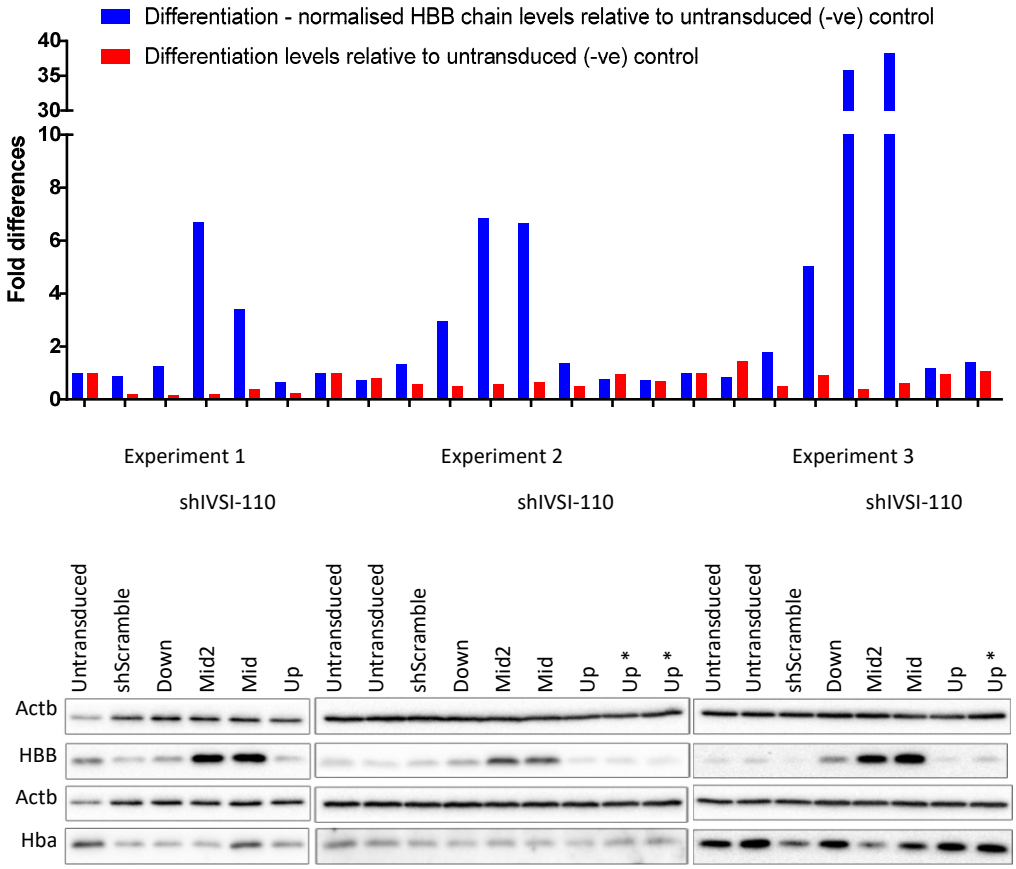


Figure 38: Immunoblots for quantification HBB and Hba chains and Actb in protein extracts from three independent transduction experiments on MEL MA821 HBB$^{IVSI-110}$ VCN 1 clone cell line at day-6 of erythroid differentiation.
Lower panel: MEL MA821 HBB$^{IVSI-110}$ VCN1 clonal cells were transduced with pLKO.I shIVSI-110 (Up, Mid, Mid2, Down) and shScramble LVs. Transduced cells were induced differentiated in induction medium (cRPMI suppl. 2% DMSO) and cells were collected for protein expression analysis on day-6 via western immunoblots analysis. Whole cell extracts from 0.5–1x10^6 cells were probed with antibodies against HBB, Hba and Actb. Actb was used as a protein loading control. Untransduced cells were used as negative control for the assessment of HBB chain expression in transduced cells. ShScramble transduced cells are used as additional negative controls. Upper panel graph: fold difference of differentiation-normalised HBB chain expression (via HBB / Hba chain ratio) and differentiation levels (via Hba chain) compared to untransduced negative (–ve) controls. ShIVSI-110 Mid and Mid2 transduced cells consistently express high levels of HBB chains in all the experiments and at lower level cells with shIVSI-110 Down. ShIVSI-110 Up and Up* designate cells transduced with MOI 10 and MOI 5, respectively. Band densities were measured using ImageLab.

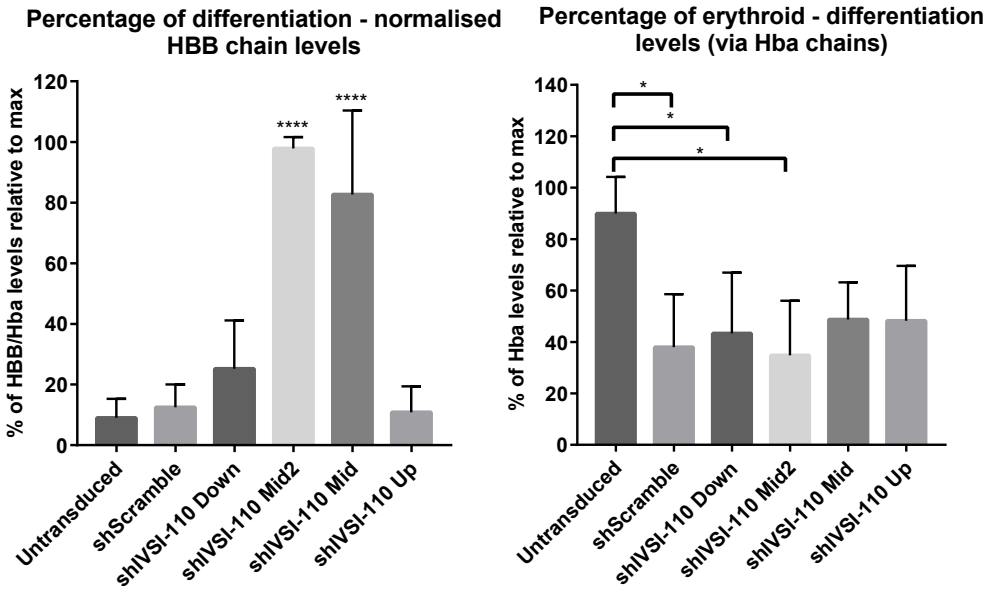


Figure 39: Percentage of erythroid differentiation (via Hba) and differentiation-normalised HBB chain levels (via HBB/Hba ratio) of three independent experiments on day-6 of differentiation, using as 100 % the sample with the highest value from each experiment.

Statistical differences of the levels of differentiation-normalised HBB chains in transduced samples were tested relative to the untransduced control, whereas for level of differentiation (determined by Hba chain levels), statistical differences were tested among all samples. Only shIVSI-110 Mid and Mid2 cells showed a significant increase in the differentiation normalised HBB chains levels relative to untransduced negative control. ShScramble transduced cells are used as an additional negative controls Note: shIVSI-110 Up* MOI 5 (n 2) is not included in the statistical analysis. Statistical analysis was performed by one-way ANOVA.

Statistical analysis:

shIVSI-110 Mid	****	HBB/Hba – *p* value 0.0001
shIVSI-110 Mid2	****	HBB/Hba – *p* value 0.0001
shIVSI-110 Mid2	*	Hba – *p* value 0.0135
shIVSI-110 Down	*	Hba – *p* value 0.0423
shScramble	*	Hba – *p* value 0.0205

* $p < 0.05$, ** $p < 0.01$*** $p < 0.001$, **** $p < 0.0001$.

3.2.6 *Assessment of translation advantage of shIVSI-110 in hCD34$^+$ cells*

Since encouraging initial data were obtained in the transgenic MEL M821 HBB$^{IVSI-110}$ VCN 1 clonal cells with the shIVSI-110 LV constructs, these vectors were then tested a series of experiments with β-thalassaemia patient-derived CD34$^+$ cells harbouring the IVSI-110 mutation. The effect of shIVSI-110 was assessed alone or in combination with the HBB LV at the translational level.

In the first instance, single transductions of pLKO.I shIVSI-110 LVs were tested on normal and HBB$^{IVSI-110}$patient-derived hCD34$^+$ cells. In general, 1x10^6 cells were thawed from liquid nitrogen storage, cultured for 24hr and then transduction conducted with 1x10^6 cells/sample.

The first experiment (pilot) was performed with the same MOIs used for the transduction of MEL MA821 HBB $^{IVSI-110}$ cells (data not shown). Assessment of VCN via multiplex qPCR allowed the

adjustment of the MOIs in order to have similar (VCN ~5) in all samples for the next experiments with normal and $HBB^{IVSI-110}$patient-derived $hCD34^+$ cells. Transductions were performed in duplicate, in which pLKO.I shScramble and an untransduced mock were also included to serve the negative controls. Cells were cultured for 48 hr before proceeding to induced erythroid differentiation. For these specific single transduction experiments cells were collected for gDNA extraction for VCN determination before culturing the cells for differentiation. The stage of differentiation was assessed on the first (data not shown) and last day of culture and analysed with Giemsa/May-Grünwald and Dianisidine staining of cells spotted on slides after cytocentrifugation (section 2.24.5). The rest of the cells were set for induced erythroid differentiation (~1 - $2x10^6$ cells) and they were cultured for 4-6 days. It was observed that during the 48 hr recovery period in expansion culture, transduced cells look healthy and continue to divide with low cell death. Conversely, when they were transferred to differentiation medium, cells are stressed and resulting in an increased cell death (Figure 40). Therefore, in general based on the state of the cultures regarding cell death (fragmented cells) and stage of differentiation (reddish cells), cultures were terminated on days 4-5. Analysis of all globin chains was assessed by HPLC analysis. Unfortunately, due to an insufficient number of cells being available on the day of collection, cells were collected only for protein extraction and HPLC analysis and not for RNA extraction. Therefore, we did not evaluate the knock-down effect of the shIVSI-110 at the level of mRNA in $hCD34^+$ cells.

Further to the single transduction experiments with all the pLKO.I shIVSI-110 RNAs LVs, it was decided to focus experiments to the most promising pLKO.I shIVSI-110 LV, Mid, alone or in combination with the MA821 HBB^{Normal} LV. Both Mid and Mid2 displayed similar results in MEL experiments and therefore the presence of the TTTT repeat did not impair the expression of the shRNA, at least not to a detectable level. Hence experiments were conducted with the unmodified version Mid. In addition, expression of the shIVI-110 RNAs in future applications on $hCD34^+$ cells would be driven by erythroid specific promoters such as an HBB promoter-LCR combination, which are used by RNA pol2. Therefore the presence of the TTTT would not interfere with the shRNA expression anyway. Panayiota Papasavva, after shared experimentation in the preceding cell culture and single-construct transduction, conducted much of the subsequent experimentation in this section under my direct supervision, including primary cell cultures, double transductions, VCN measurements and scoring of cytocentrifugation samples and *in silico* analysis of HPLC data, towards receipt of an MSc from the Cyprus School of Molecular Medicine.

In all experiments transduction with pLKO.I shScramble LV and untransduced negative control samples were included. Single and double transductions were done in parallel. In particular, initially

cells destined for double transductions, were transduced all together with MA821 HBB^{Normal} LVs on day 1, whereas those destined for single transduction were kept in culture until day 2. Transductions with the MA821 HBB^{Normal} LV were performed on day 1 in order to obtain the same *HBB* transgene background in all the cells prior to the addition of the second LV. On day 2 both, single and double, transductions with the pLKO.I shIVSI-110 Mid and shScramble were completed. Since the number of transduced samples was narrowed to shIVSI-110 Mid and shScramble, we could perform duplicates per experiment, which increased the accuracy and assessed reproducibility of results.

3.2.6.1 *Vector copy number of transduced CD34^{+} cells*

Quantification of the average VCN/cell in transduced samples was determined by a multiplex qPCR method [235]. As mentioned above, in the first single transduction experiments with all the pLKO.I shIVSI-110 LVs, Up, Mid, Mid2, and Down, cells for VCN quantification were collected prior to erythroid differentiation of cultures (48-post-transduction), whereas in the latter single and double transduction experiments with the pLKO.I shIVSI-110 Mid and shScramble, cells were collected on the last day of differentiation. This gave more accurate and reliable results, since increased cell death was observed when transduced cells were transferred for induced differentiation. This could lead to an overestimation of the average VCN, since the constitution of the population changes. In addition, non-incorporated LVs could also be included in the final VCN quantification, although it does not actually contribute to the expression of the shRNAs and *HBB* transgene.

In general, at 48hr post-transduction cells displayed on average ~ 5% cell death (assessed by trypan blue staining), whereas on the last day of differentiation the average cell death was ~ 43%. On the contrary, cell death in the untransduced cells changed from 2.6% to 16% and in samples transduced only with MA821 HBB^{Normal} LV ranged from 2.8% to 12.73% (Figure 40 & Figure 42).

On one hand, quantification of samples from cells transduced with the pLKO.I shIVSI-110 and shScramble exhibited very high VCN of ~25.5. Whilst on the other hand, cells transduced with the MA821 HBB^{Normal} LV showed an average VCN ~ 2.69 (Figure 41 & Figure 42). The increased VCN found in cells transduced with pLKO.I shIVSI-110 and shScramble could explain the elevated levels of cell death on the last day of differentiation compared to the untransduced or transduced cells with only MA821 HBB^{Normal} LV. High VCN is associated with high cell death due to increased cytotoxicity, probably caused by off-target effects from the high expression of shRNA, increased genome instability and disruption of vital genes [245]. In addition, increased cell death was observed mainly only when cells were transferred to the differentiation medium. Alteration of the culture medium

might have stressed the cells, which in combination with the high VCN rendered them even more permissive to trigger apoptosis.

Samples in which the VCN quantification was assessed on the last day of differentiation, are characterised by lower and more consistent VCNs among duplicates within the same experiment, compared to the samples were the VCN was assessed on day-0 of differentiation (Figure 41).

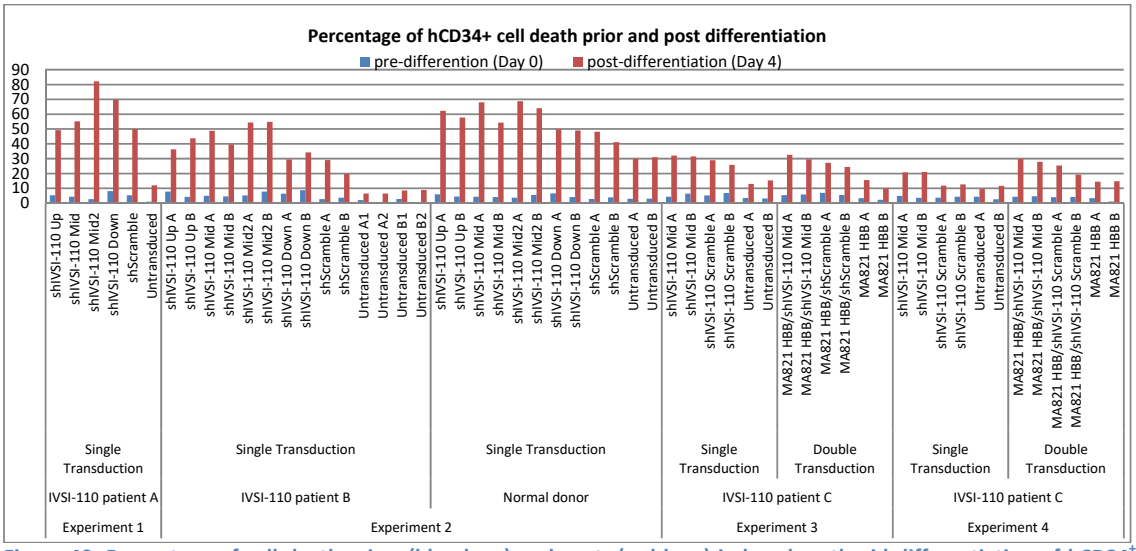


Figure 40: Percentage of cell death prior- (blue bars) and post- (red bars) induced erythroid differentiation of hCD34$^+$ cells in transduction experiments.

Experiments were performed of hCD34$^+$ cells from three different patients (A, B and C) and normal donor. Duplicates within the same experiments are designated as A and B, and separate differentiation cultures derived from the same sample A or B are designated as A1 and A2 or B1 and B2, respectively. Cells death was measured microscopically with trypan blue stain.

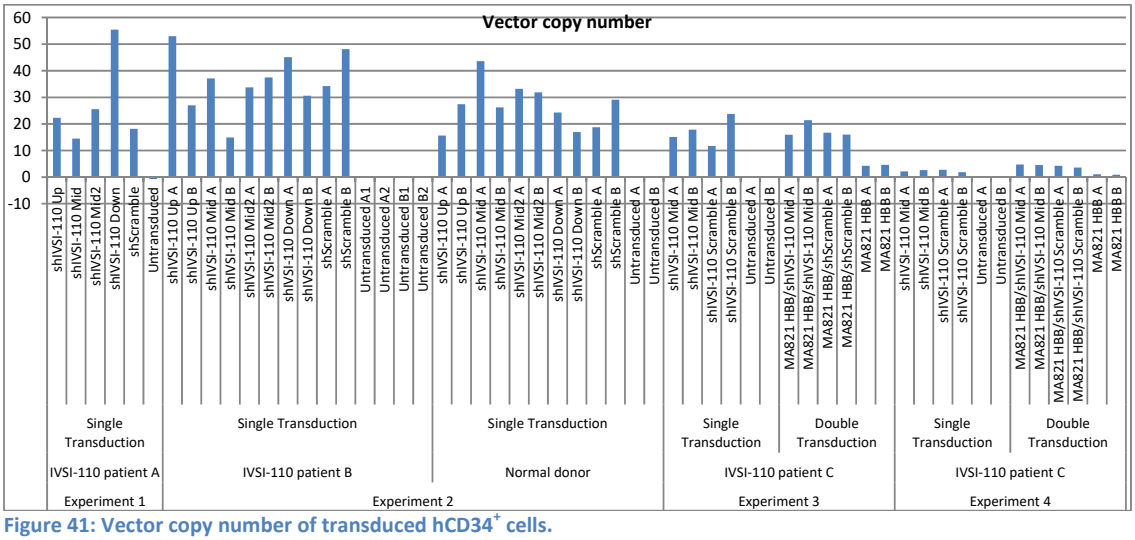


Figure 41: Vector copy number of transduced hCD34$^+$ cells.

Average VCN/cell in the transduction experiments with all pLKO.I shIVSI-110 LVs (Up, Mid, Mid2, Down) was measured before induction of differentiation, which is a stage characterised with relatively low cell death. VCN in the transduction experiments focused on the pLKO.I shIVSI-110 Mid, the VCNs were measured on the last day of differentiation, which is characterized by increased cell death. In the latter case the VCN values are lower compared to the other experiments and more consistent between duplicates. Experiments were performed of hCD34$^+$ cells from three different patients (A, B and C) and normal donor. Duplicates within the same experiments are designated as A and B, and separate differentiation cultures derived from the same sample A or B are designated as A1 and A2 or B1 and B2, respectively. VCN was measured via a multiplex qPCR method.

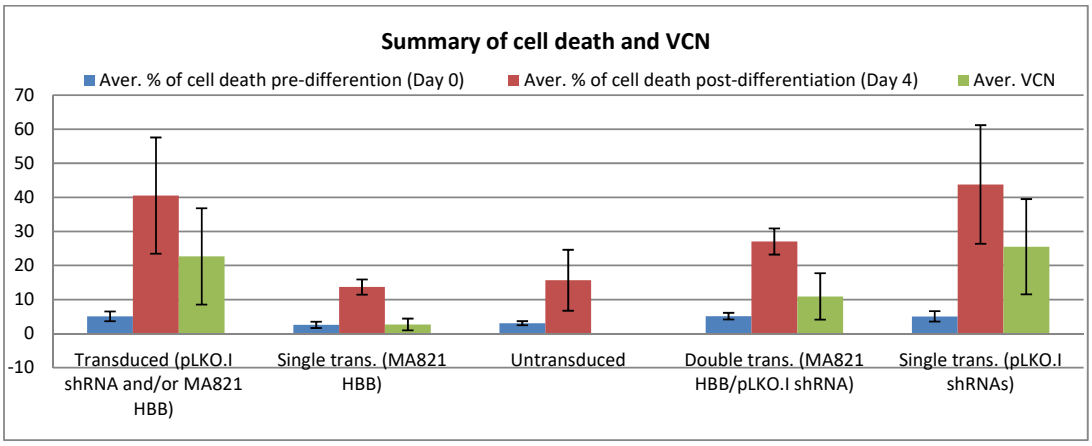


Figure 42: Summary of average cell death prior- (blue bars) and post- (red bars) induced erythroid differentiation of hCD34^{+} cells and average VCN (green bars) in all transduction experiments.
Samples are categorised in transduced (pLKO.I shRNA and/or MA821 HBB^{Normal} LVs), single transduced (MA821 HBB^{Normal} or pLKO.I shRNA LVs), double transduced (pLKO.I shRNA and MA821 HBB^{Normal} LVs) and untransduced cells.

3.2.6.2 *Stage of differentiation (May Grünwald / Giemsa / Dianisidine staining)*

The stage of erythroid differentiation of hCD34^{+} samples was assessed on the last day of differentiation (day 4) based on the morphology and level Hb expression, by staining for May-Grünwald/Giemsa stain and dianisidine, respectively, as described in section 2.24.5. Images from different fields of the stained cells were counted and characterised based on their morphology as ProEr, Baso, Poly, Ortho and reticulocytes (Ret) and as dianisidine/Hb positives cells. Data from all the experiments were put together and are given as the percentage of each category relative to the total population. In Figure 43 examples of the different stages of erythroid differentiation are shown. It is noteworthy the higher proportion of cells at the orthochromatophilic and reticulocyte stage as also the higher percentage of Hb expressing cells (dianisidine positives) in the normal donor sample relative to the thalassaemic untransduced control, where cells are at more immature stages with lower percentages of dianisidine positives . This is expected, since normal amounts of HBB chains are synthesised in the healthy donor and their erythropoiesis is not affected. In a previous study, they showed that the erythroid differentiation of thalassaemic cells is blocked and in particular they demonstrated a reduced progression of the cells into the orthochromatic normoblast stage and increased levels of cell death at the polychromatophilic stage relative to the normal donor or samples transduced with the GLOBE LV [115].

Overall, differential counting (Figure 44) indicated a significant increase of the percentage of cells at late stage erythropoiesis (Ortho+Ret) in samples transduced with pLKO.I shIVSI-110 Up, Mid and Mid2 and in double transduced with MA821 HBB^{Normal}/shIVSI-110 Mid. Differential counts on normal donor-derived CD34^{+} cells, as expected indicated a significant increase of the late stage

erythropoiesis relative to the counts on untransduced HBB$^{IVSI-110}$patient-derived cells, whereas similar percentage of Hb expressing positive cells were identified (Figure 44).

Overall, these data comprises encouraging indices for correction of the thalassaemic phenotype in cultures expressing the shIVSI-110 RNAs, and especially with shIVSI-110 Mid and Mid2, in which cultures are in mature stage of erythroid differentiation (Ortho+Ret) relative to the negative controls (shScramble and untransduced cells).

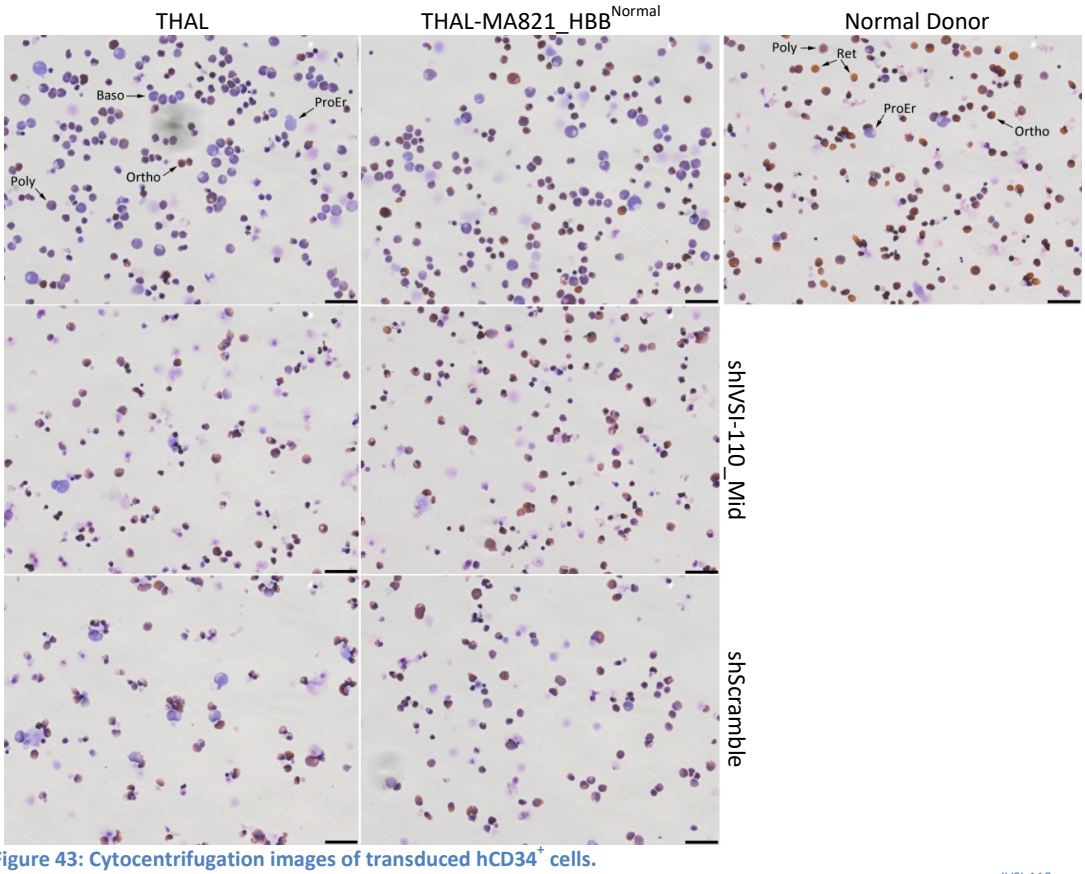
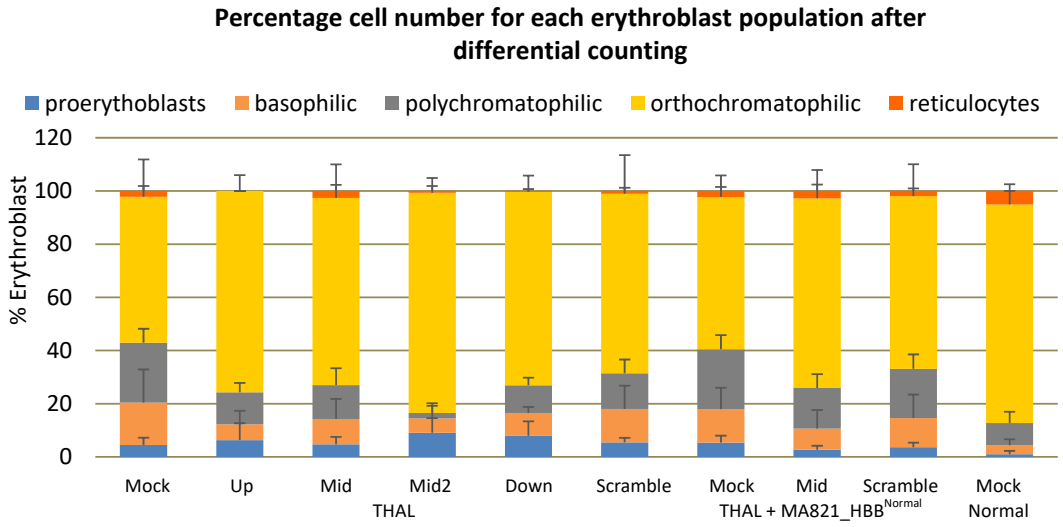


Figure 43: Cytocentrifugation images of transduced hCD34$^+$ cells.
Representative images of cytocentrifugation samples, stained with May-Grünwald–Giemsa–Dianisidine of HBB$^{IVSI-110}$patient-derived- (THAL), transduced MA821 HBB^{Normal} HBB$^{IVSI-110}$patient-derived- (THAL-MA821 HBB^{Normal}) and untransduced normal-donor derived- (ND) CD34$^+$ cells on day 4 of induced erythroid differentiation. THAL and THAL-MA821 HBB^{Normal} shown were further transduced with pLKO.I shIVSI-110 Mid or shScramble LVs. Different stages of cell erythroid differentiation are designated with black arrows in the untransduced samples from thalassaemic$^{IVSI-110}$ (THAL) and normal donor (ND) cells. Dianisidine positives are stained brown. Proerythroblast, ProEr; Basophilic, Baso; Polychromatophilic, Poly; Orthochromatophilic, Ortho; Reticulocytes, Ret. Magnification 40 X, — Scale bar: 20 μm.


Percentage cell number for each erythroblast population after differential counting
proerythoblasts
basophilic
polychromatophilic
orthochromatophilic
reticulocytes
% Erythroblast
120
100
80
60
40
20
0
Mock
Up
Mid
Mid2
Down
Scramble
Mock
Mid
Scramble
Mock
THAL
THAL + MA821_HBB Normal
Normal

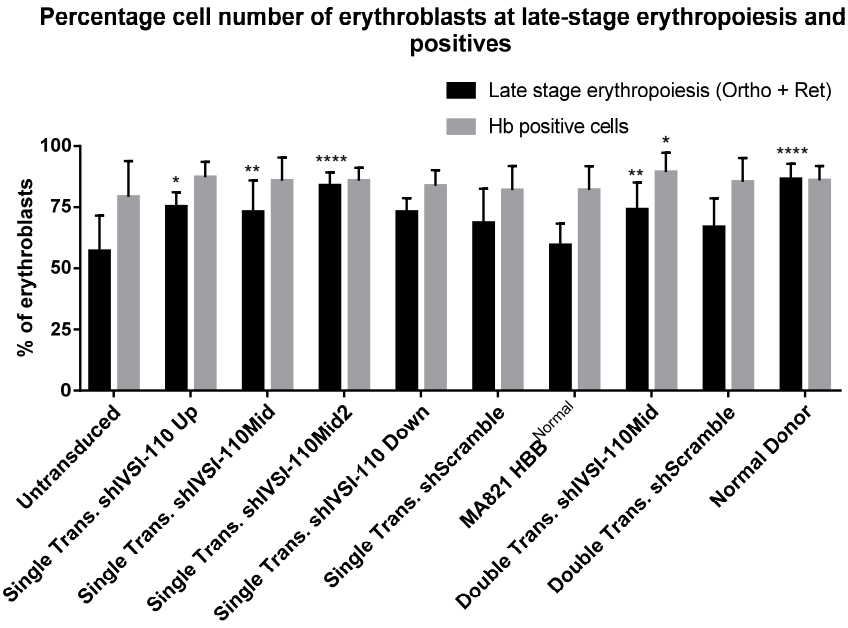

Percentage cell number of erythroblasts at late-stage erythropoiesis and Hb positives
Late stage erythropoiesis (Ortho + Ret)
Hb positive cells
% of erythroblasts
100
75
50
25
0
*
**
****
**
*
****
Untransduced
Single Trans. shIVSI-110 Up
Single Trans. shIVSI-110Mid
Single Trans. shIVSI-110Mid2
Single Trans. shIVSI-110 Down
Single Trans. shScramble
MA821 HBB Normal
Double Trans. shIVSI-110Mid
Double Trans. shScramble
Normal Donor

Figure 44: Differential counting of transduced patient-derived CD34^{+} cells.

(Top panel) Percentage cell number for each erythroblasts population after differential counting of May- Grünwald - Giemsa-Dianisidine stained cytocentrifugation samples from day-4 erythroid differentiated cultures of β-thalassaemia HBB$^{IVSI-110}$patient-derived CD34^{+} (n=3) (THAL). THAL cells were transduced with pLKO.I_shIVSI-110 (Up, Mid, Mid2, Down) and shScramble LVs or double transduced with MA821_HBB^{Normal} LV and pLKO.I_shIVSI-110_Mid or shScramble. Untransduced (mock) samples serve the negative controls, whereas counts of untransduced (Mock) normal donor serve the positive control.

(Bottom panel) Percentage of late stage erythropoiesis and Hb expressing cells (Dianisidine positive cells). Statistical differences of the percentage of late stage erythropoiesis and Hb expressing cells were tested relative to the untransduced negative control. Patient-derived CD34^{+} cells transduced only with pLKO.I shIVSI-110 Up, Mid, Mid2 (single trans.) and pLKO.I shIVSI-110 Mid in combination with MA821 HBB^{Normal} (double trans.) indicated a statistical significant increase of the percentage of cells in the late-stage erythropoiesis (orthochromatophilic and Reticulocytes, black bars) relative to the untransduced cells. The double trans. shIVSI-110 Mid exhibit a significant increase of percentage of cells expressing haemoglobin (Hb) (dianisidine positive) compared to the untransduced. Percentage of cells in the late stage of erythropoiesis was significantly higher in normal donor samples compared to the patient-derived CD34^{+} untransduced cells. Increase of the percentage of late stage of erythropoiesis is associated with correction of β-thalassaemia phenotype. Statistical analysis was performed by one-way ANOVA.

Statistical analysis

Untransduced control compared to:

shIVSI-110 Up	*	Ortho+Ret	– *p* value 0.0290
shIVSI-110 Mid	****	Ortho+Ret	– *p* value 0.0056
shIVSI-110 Mid2	****	Ortho+Ret	– *p* value <0.0001
MA821 HBB/^{I}shIVSI-110 Mid	**	Ortho+Ret	– *p* value 0.0085
Normal donor	****	Ortho+Ret	– *p* value <0.0001
Double Trans.^{I}shIVSI-110 Mid	*	Hb +ve	– *p* value 0.0227

* $p < 0.05$, ** $p < 0.01$*** $p < 0.001$, **** $p < 0.0001$.

3.2.6.3 *Effect of shIVSI-110 RNAs on HBB protein levels in transduced patient-derived CD34$^+$ cells.*

Our aim is to increase the therapeutic potential of *HBB*-based LVs, by reducing the competition between the two HBB variants for translation factors via knocking-down specifically the relatively stable aberrant HBB transcript. Transduced cells were harvested on days 4–6 post-erythroid differentiation for protein extraction and analysed by HPLC for HBB chains. In general, approximately 1x10^6 differentiated cells are sufficient for HPLC analysis. The HBB expression levels are given as the ratio of the peak areas of HBB to HBA chains, whereas comparisons are determined as the percentage change of the ratios relative to the untransduced (negative) control.

Analysis of protein extracts from differentiated patient-derived CD34$^+$ cells transduced with only pLKO.I shIVSI-110 shRNAs showed a significant increase in the percentage HBB/HBA chain ratio compared to the untransduced control from 31.87±16.09%, 44.00±20.57% and 29.28±26.54% in samples transduced with the shIVS-110 Mid (n=7, *p* value = 0.0031), Mid2 (n=3, *p* value = 0.0019) and Down (n=3, *p* value = 0.041), respectively. A marginal increase was also detected in samples transduced with shIVSI-110 Up and shScramble, by 10.39±8.69% (n=3) and 8.18±10.20% (n=3) compared to the untransduced 0.00±4.00 (n=9) samples respectively, albeit not statistically significantly (Figure 46).

Since in all the MEL experiments the shIVSI-110 Mid and Mid2 constructs showed similar results, in all probability the TTTT in the sense sequence strand of MID does not negatively affect its expression. Taking in account all the above, two independent experiments with single transduction and double transduction of patient-derived CD34$^+$, with pLKO.I_shIVSI-110 Mid alone, or in combination with MA821 was undertaken. Theoretically, addition of the MA821 HBB^{Normal} will increase the completion for translational factors making the effect more obvious as also it will provide data regarding the level of optimisation of therapeutic potential of the MA821 HBB^{Normal} LV.

In the case of double transductions, where cells were transduced initially with the MA821 HBB^{Normal} LV and subsequently with the shIVSI-110 Mid and shScramble LVs, a significant increase in HBB/HBA was observed in the samples transduced with the shIVSI-110 Mid, from 27.2±15.63% (n=4) of MA821 HBB background to 59±15.65% (n=4; p value 0.003). Double transduction with shScramble also showed higher levels of the HBB chains with a 35.4±23.52% (n=4) change being observed compared to untransduced samples; this is higher from the levels obtained with MA821 HBB^{Normal} only, but lower compared to the transductions with shIVSI-110 Mid (Figure 46).

One of the most important observations was made from the comparison of data obtained from the single transductions of shIVSI-110 Mid, Mid2 and Down and MA821 HBB^{Normal}. In this case the levels of the HBB chains in samples transduced with only the shIVSI-110 Mid (31.87±16.1%; n=7), Mid2 (44±20.67%; n=3) and Down (29.28±26.54%; n=3) reached or even exceed the HBB chain levels (27.2±15.63%; n=4) obtained in samples containing the MA821 HBB^{Normal} LV at an average VCN ~2.7 in patient-derived $CD34^+$ cells. The HBB chains levels in cells transduced with shScramble LV remained close to background levels (8.18±10.2%; n=7) (Figure 46). According to the results from single and double transductions, addition shIVSI-110 Mid/Mid2 RNAs, is leading to an increase of the HBB chains by 30-44% of the background levels (endogenously or/and vector-derived).

HPLC analysis from a single experiment on a normal sample showed an average decrease in the ratio of HBB/HBA relative to the untransduced controls, only in the case of the shIVSI-110 Down shRNA by 9±2% (n=2) and shScramble by 2.8±3.9% (n=2). In contrast, the remaining samples exceeded the normal levels in particular, by 15±4% (n=2), 24±5.2% (n=2), 46±34.75% (n=2) in cells transduced with shIVSI-110 Up, Mid and Mid2, respectively. These preliminary results on normal $CD34^+$ cells suggest that our shIVSI-110 RNAs with the exception of the shIVSI-110 Down construct are not targeting the correctly spliced mRNA, at least not at a degree that would result in an observed decrease in HBB chains below normal levels. On the contrary, an increase in HBB chains is detected, even in cells isolated from normal donors, where no shIVSI-110 RNA targets would be expected and where the only available target for shIVSI-110 RNAs would be incompletely processed pre-mRNAs or aberrant transcripts derived from the endogenous *HBB* gene. Abundant, chimeric aberrant HBB transcripts were reported for lentiviral integration of the GLOBE-encoded *HBB* [246], and it has been shown that alternative splicing coupled with NMD is used by cells to regulate gene expression post-transcriptionally [247]. Interpreting the data in Figure 45 in this light and postulating accelerated clearance of naturally occurring aberrant transcripts by action of shIVSI-110 RNAs is therefore tempting. However, these data are based on a single experiment and for the normal sample do not tie in with publications in the field, which presently provide no prior reports of abundant and potentially interfering aberrant transcripts for the normal *HBB* endogene. In the absence of further data, interpretation of apparent shIVSI-110-RNA-enhanced HBB expression in normal samples is therefore unprofitable.

Importantly, preliminary data on thalassaemic $CD34^+$ cells showed the first encouraging data, especially in the case of Mid and Mid2, which support our hypothesis for competition between the two HBB variants for translation factors and are in agreement with the data obtained in MEL MA821 $HBB^{IVSI-110}$ cells, although the level of increase is not comparable. This could be due to fact that the

competition in the murine system is higher because of additional expression of the endogenous murine *Hbb-b1/b2*. In the cases where increased competition between the two variants was attempted by co-transduction of the shIVSI-110 Mid and MA821 HBB LV, the level of HBB chains increased the same with the increase seen without the MA821 HBB^{Normal} (by ~ 32%) Figure 46. According to these data, the shIVSI-110 Mid is able to significantly improve the therapeutic potential of MA821 HBB^{Normal} LVs by ~30% when used in HBB$^{IVSI-110}$-patient-derived CD34^{+} cells, which would probably allow IVSI-110 patients to reach to transfusion independence at lower VCN constituting the GT application safer. However, it should be noted that this was observed with MA821 HBB^{Normal} VCN ~2.7 whereas in the clinical objective is an average VCN 0.2-2 for safety reasons. It was also noticed that in liquid cultures of patient-derived CD34^{+} cells the ratio of HBB/HBA chains are high (0.41) compared to what could be expected in thalassaemic patients (0.08-0.2). Therefore, although these data are encouraging further experiments should be performed, more importantly with lower average VCN/cell for both constructs that will assist in better understanding the actual effect of the LV tested on thalassaemic CD34^{+} cells.

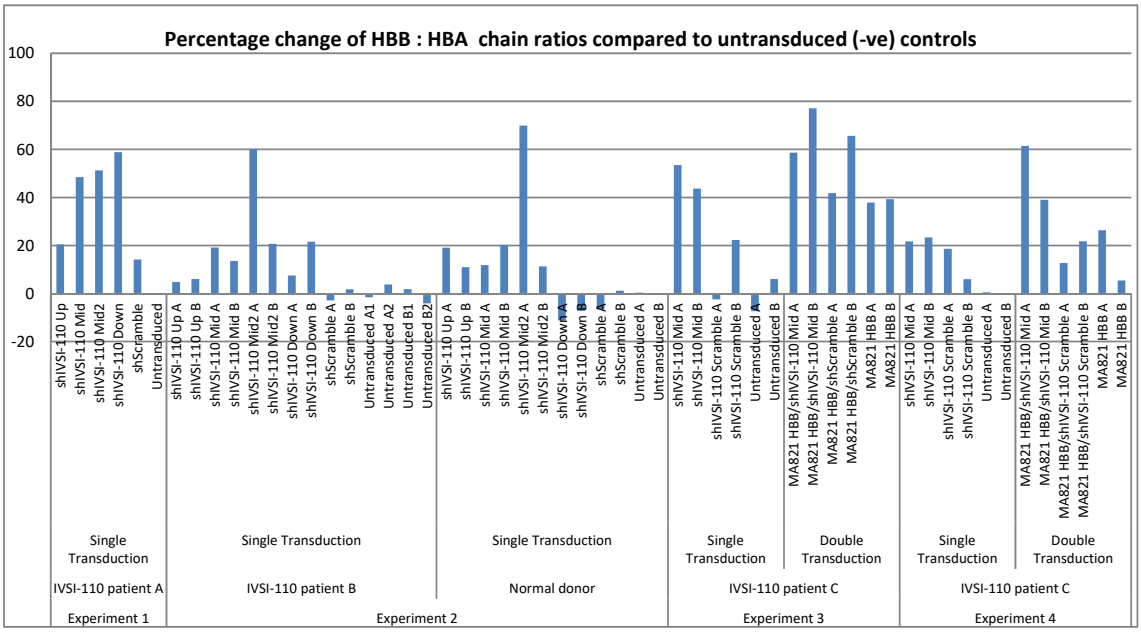


Figure 45: Percentage change of HBB:HBA chain ratio of transduced hCD34^{+} cells relative to the untransduced.
Percentage change of the HBB:HBA chain ratios, measured by HPLC, of transduced hCD34^{+} cells compared to the untransduced negative (-ve) controls in each experiment (experiments 1–4) after 4–5 days of induced erythroid differentiation. In experiments 1–2 all versions of the shIVSI-110 RNAs were tested on A and B patient- or normal- derived hCD34^{+} cells. In experiment 3–4 single and double transduction were performed on patient C derived hCD34^{+} cells. Duplicates within the same experiments are designated as A and B, and separate differentiation cultures derived from the same sample A or B are designated as A1 and A2 or B1 and B2, respectively.

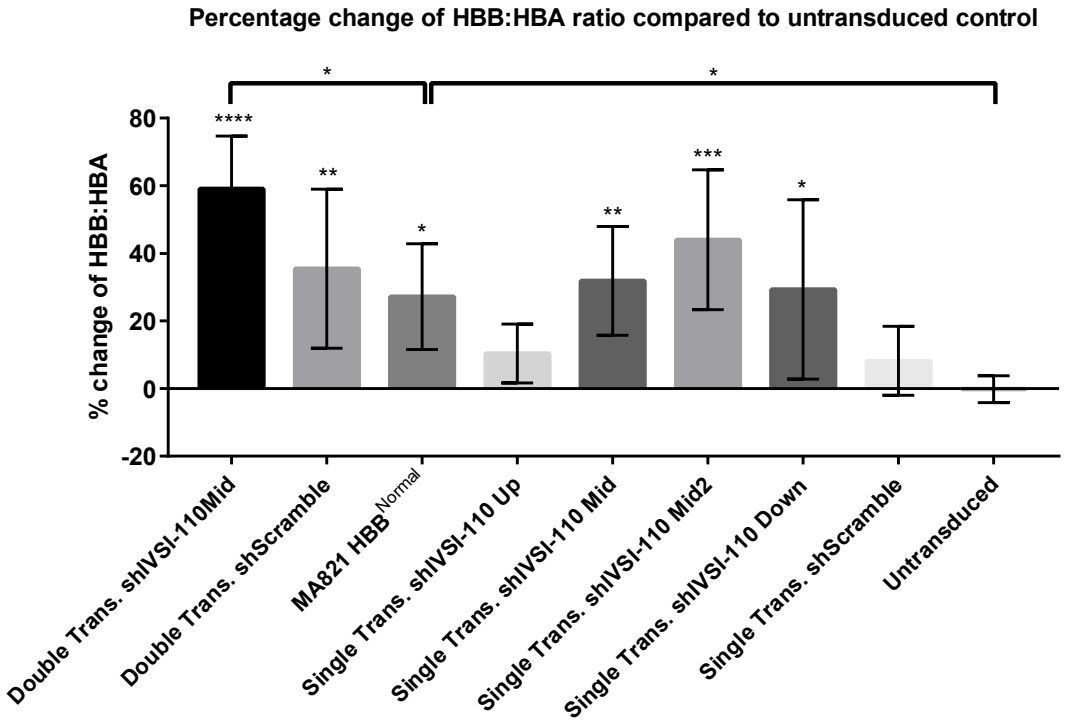


Figure 46: Assessment of translational advantage of shIVSI-110 RNAs on thalassaemic CD34^{+} cells at a protein level. Representation of the percentage (%) change in the HBB:HBA chain ratio in transduced cells from all experiments in patient-derived cells relative to the untransduced control. In this set of experiments, in addition to the single transductions, cells were also transduced with the MA821 HBB^{Normal} LV, (double transductions, experiment 3–4, duplicates). Statistical differences were tested relative with untransduced sample and relative to the single transduced MA821 HBB^{Normal} cells. Statistical analysis was performed by one-way ANOVA.

Statistical analysis:

Untransduced control compared to:

MA821 HBB / shIVSI-110 Mid	****	– *p* value 0.0001
MA821 HBB/ shScramble	**	– *p value* 0.0025
MA821 HBB^{Normal}	*	– *p value* 0.0276
shIVSI-110 Mid	**	– *p* value 0.0011
shIVSI-110 Mid2	***	– *p* value 0.0007
shIVSI-110 Down	*	– *p* value 0.0344

MA821 HBB^{Normal} control compared to:

MA821 HBB / shIVSI-110 Mid	*	– p value 0.003
Untransduced control	*	– p value 0.0276

* $p < 0.05$, ** $p < 0.01$*** $p < 0.001$, **** $p < 0.0001$.

3.2.7 *Conclusions: Gene addition*

In this section, the development of a novel gene addition approach for personalised GT for β-thalassaemia mutations that lead to persistent and non-functional mRNA, such the HBB$^{IVSI-110(G>A)}$ has been described. HBB$^{IVSI-110 (G>A)}$ is a misspliced mutation, which leads mainly to the production of an aberrant HBB variant mRNA containing 19 nt of intron-1 sequences and the formation of a pre-mature termination codon (PTC) recognised in turn by the NMD pathway with aberrant mRNA targeted for destruction [86]. Although, the majority of pre-mRNA is aberrantly spliced, a fraction of

pre-mRNA is correctly spliced. Therefore functional HBB chains are still synthesised at ~10% of the amount that are produced in healthy individuals [90].

We believe that the presence of the aberrant mRNA competes with the correctly spliced mRNA, endogenously- or vector- derived, for translation factors and thus compromises normal HBB chain production. Our hypothesis is based on preliminary data from the direct comparison of the expression level from the T9W *HBB* lentiviral vectors with and without ankyrin insulator elements in patient-derived $CD34^+$ cells with β^0 and β^+ thalassaemia [222]. Similar to the findings presented here these authors showed that $HBB^{IVSI-110}$ aberrant mRNA is more stable than HBB^{cd39} mRNA and that it consists on average 49% of total HBB mRNA. They showed that whereas HbA levels in differentiated T9W-tansduced β^0 $CD34^+$ cells increased in a VCN dependent manner, T9W-tansduced $CD34^+$ cells from $\beta^{+/+}$ and $\beta^{0/+}$ patients produced little or no HBB chains, even though high levels of transgenic HBB mRNA was detected implying the presence of competition between two HBB variants for translational factors. This competition for translation was overcome with the addition of the Ankyrin insulator element on T9W, AnkT9W, which allowed the high expression of the HBB transgene at earlier stages of erythroid differentiation [222].

In addition, if the proposed hypothesis is valid and the presence of this competition has a direct impact on the amount of HBB chains that is synthesised, this limits the therapeutic potential of *HBB*-based LVs, such the MA821 HBB^{Normal} LV (GLOBE) [115]. Consequently, this means that in such instances LV-corrected cells may need to present at higher numbers in order to become therapeutic, increasing at the same time the risk for insertional mutagenesis, activation of proto-oncogenes and disruption of other gene functions. In addition, the requirement of a larger amount of LVs also increases the cost of an already expensive treatment per patient.

The experiments described in this section present a strategy to test the hypothesis of competition of $HBB^{IVSI-110}$ mutant mRNA for translational machinery. The strategy presented involves removal of the competition between the two HBB mRNA variants (IVSI-110 and wild-type) with the use of short hairpin RNAs specific for targeting the aberrant mRNA. A total of four shIVSI-110 specific RNA (Up, Mid, Mid2 and Down) were designed and produced and expressed from within pLKO.I LV vectors.

The initial aim was the simultaneous expression of both wild-type *HBB* transgene and shIVSI-110 RNAs from the same construct, in which the shRNA was cloned at a specific site in Intron-II of the *HBB* transgene. Therefore, with the splicing of the HBB pre-mRNA the shRNA would be released with intron-2 and proceed with knock-down of the aberrant mRNA. Although, all MA821 HBB shIVSI-110 RNA LV versions were produced and titrated, a parallel work in which the same vector design strategy was followed for targeting BCL11A transcripts, showed that the shRNA in not active even-

though the transgene was expressed. This implied that the cloning site and context around the shRNA is not suitable to be recognized and processed accordingly by DICER.

Therefore, before proceeding to a more complicated delivery and expression system, proof of principle was attempted by delivering and expressing the IVSI-110 mutant mRNA specific shIVSI-110 RNAs (Up, Mid, Mid2 and Down) with the widely used pLKO.I LV, which constitutively express the shRNA under the U6 promoter. The results obtained establish proof of principle in a transgenic MEL $HBB^{IVSI\text{-}110(G>A)}$ clonal cell line with VCN 1 in which generally expresses approximately 4.2% of the HBB chains compared to MEL MA821 HBB^{Normal} control cells with an average VCN of 2. The data obtained show a significant increase in HBB chain synthesis by 8-10-fold than the untransduced samples with the shIVSI-110 Mid and Mid2 RNAs only. However, knock-down of the aberrant mRNA fraction was not observed which implies that maybe shIVSI-110 Mid/Mid2 increase the translation rate of the correctly spliced HBB mRNA, rather via translation repression of than RNA degradation of the aberrant mRNA. Translational advantages were also observed in experiments with patient-derived $CD34^+$ cells. One of the most impressive outcomes from these experiments came from the comparison of the HBB/HBA chain ratio between differentiated $hCD34^+$ cells transduced with the shIVSI-110 MID LV and those with MA821 HBB^{Normal} VCN ~2.7. Surprisingly, cells expressing only the shIVSI-110 Mid shRNA reached and exceeded the amount of HBB chains that are synthesised in cells expressing on average ~2.7 HBB^{Normal} transgenes. In general, we showed that targeting the aberrant HBB mRNA with shIVSI-110 RNA, and in particular Mid and Mid2, is able to increase the levels of HBB/HBA chain ratio by 30-44% of background levels. Although no knock-down effect was detected in MEL experiments with the Mid and Mid2, we believe that the increase of the HBB chain synthesis is the result of reducing the competition at a point where the availability of translation factors for the correctly spliced mRNA is increased. However further experiments need to be performed for the better understanding of the mechanism responsible for the increased translation of normal HBB mRNA as also the clarification of silencing pathway which shIVSI-110 Mid follows; RNA degradation or translation repression.

3.3 Genome editing approach

3.3.1 *Aims:*

Development and *HBB*$^{IVSI-110}$ specific designer nucleases:

Transcription like effector nucleases (TALENs)

RNA guided endonucleases (RGEN)

Assessment of comparative targeted disruption efficiencies:

TALEN pair Vs RGEN

Modified versions of TALEN pairs

Assessment of functional correction of the *HBB*$^{IVSI-110(G>A)}$ in:

Humanised transgenic MEL MA821$^{IVSI-110(G>A)}$ cell line

Patient-derived hCD34^{+} cells

3.3.2 *Experimental strategy*

A described above the *HBB*$^{IVSI-110(G>A)}$ mutation creates an alternative splicing SA site (AG) 19 nucleotides upstream from the normal AG of IVSI. The 19nt are abnormally spliced into the mature mRNA, changes the ORF creating a PTC which is recognized by the NMD pathway and leads to mRNA degradation. Therefore no functional HBB chain is produced. The original SA site is still used by the spliceosome but at a low frequency (10-20%), which means that there is some production of functional HBB chains but is not sufficient to provide a transfusion independent thalassaemia intermedia phenotype [82,90,241].

Hypothesis

It is hypothesised that disruption of the alternative SA site (AG), produced by the *HBB*$^{IVSI-110(G>A)}$ mutation will force the spliceosome to use the normal SA site, and therefore restore the correct splicing of HBB pre-mRNA. In order to achieve this, a number of genome editing nucleases specific for the region around HBB$^{IVSI-110(G>A)}$ mutation. These genome editing tools include a single RGEN specific to create a DSB on the mutation and several TALEN pairs, which theoretically creates a DSB upstream of the mutation. Targeted DSBs can then be repaired by the HDR or NHEJ repair pathways. Our approach relies on the correction of the DSBs by the efficient and error-prone NHEJ repair pathway, which produces INDELs at the site of the break [168]. The INDELs occur in close proximity to the mutation and as a result may be able to disrupt the context of the alternative SA site caused by the HBB$^{IVSI-110(G>A)}$ mutation and the adjacent region, in way that will favour the use of the normal SA site for splicing.

Up to now the gene correction strategy that is been followed by most of the groups, is through the HDR pathway with the aim to complete accurate correction of mutations, in which a donor DNA template with the wild type sequence is provided as template to be used for HR. As mentioned in the introduction, the HDR pathway occurs at low frequency, and therefore clonal selection is required in order to select for corrected cells. However, clonal selection is not applicable to $hCD34^+$, the target cell population for thalassaemia GT [39], due to their limited lifespan in culture.

Based on bone marrow (BM) transplantation studies it has been shown that 20% chimerism of the BM with normal $hCD34^+$ cells is sufficient to provide adequate therapeutic benefit to render a thalassaemic patient transfusion independent. This is mainly due to the full differentiation potential of normal or corrected $hCD34^+$ and the survival advantage of normal RBC over their thalassaemic counterparts. This makes the therapeutic approach presented here more feasible, since in principle only a fraction of corrected cells is required to rescue the thalassaemic phenotype in patients.

3.3.3 *Design and production of $HBB^{IVSI-110}$ specific designer nucleases*

Based on the nature of the gDNA sequence adjacent to the $HBB^{IVSI-110(G>A)}$ mutation, it was possible to design a single RGEN, and three TALEN monomers, two left (L1 and L2) and one right (R1).

3.3.3.1 *RGEN*

Briefly, the CRISPR/Ca9 genome editing system has two components: a) the gRNA, which is an engineered fusion RNA molecule derived from the two RNA molecules originally found in bacteria, the CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA), and b) the non-specific CRISPR-associated endonuclease (Cas9). The gRNA contains the scaffold sequence, which is required for binding to Cas9 and the target sequence (20 nt), which is necessary for the RGEN specificity to DNA target sequence. The specificity of the RGEN system is driven by Watson and Crick base pairing between the 20 nt guide sequence on the gRNA and target genomic sequences (gDNA), and the presence of the Cas9-specific protospacer adjacent motif sequence, also known as the PAM sequence, immediately downstream of the target sequence. Different Cas9 enzymes have different PAM sequences, with the Cas9 from *Streptococcus pyogenes* and its PAM sequence of 5'-NGG-3' being used in this study. CRISPR/Cas9 complex initially starts to bind to all the sites with the appropriate PAM sequence and if the guide sequence of the gRNA also matches with the binding sequence, a DSB is produced, 3-4 bases upstream from the PAM sequence. According to the sequence around the $HBB^{IVSI-110(G>A)}$ mutation, only a single suitable site was found with the necessary PAM sequence, which ultimately will lead to production of a DSB directly on the site of the mutation (Figure 47).

3.3.3.2 *TALENS*

Transcription activator-like effector nucleases (TALENs) are specific endonucleases and they were engineered by fusing the TALE's DNA binding domain with the endonuclease domain of the FOKI restriction enzyme. The TALE DNA binding domain is composed of 13–28 repeat monomers of 33–35 amino acids. Each monomer is able to recognise and bind a single nucleotide. The amino acid sequence of each monomer is highly conserved with the exception of the 12th and 13th amino acids, which are also known as repeat variable di-residues (RVD). The specificity of each monomer is determined by the RVD, for example RVD: NG recognises T, HD recognises C, NI recognises A, NN recognises G and A, and NK recognises G. Monomers on the TALEs DNA binding domain are arranged accordingly, in order to bind on a specific sequence on the gDNA. The *Fok*I endonuclease domain is able to cleave gDNA when it is dimerized with another *Fok*I. Therefore TALENs act as pairs, which in combination with the larger DNA binding target sequence provide increased specificity for the target sequence and low off-target potential, both essential properties for designer nucleases.

The Golden Gate assembly platform [227] was employed for the production of the *HBB* specific TALEN constructs. The assembly method is based on Type SII restriction enzyme sites, *Bsa*I, which the cleavage of DNA occurs at a different location than the DNA restriction site. Therefore, monomer constructs from the kit plasmid library are selected based on their position on the TALEN and digestion with *Bsa*I, modules containing the desired RVD are released from the constructs with unique overhangs, which lead to an ordered ligation on monomers to specific level II vectors forming the three intermediate multimers (1–3). Similarly, in the second step the three multimers are released by the digestion with *Bpi*I and the unique overhangs allow the formation of the full TALE repeat in the right order in the final TALEN construct (Level III). The final Level III vector includes the final half-TALE repeat (17.5), CMV promoter, HA-taq, NLS, N' and C' terminal of TALE repeat, the wild type *Fok*I endonuclease cleavage domain and the poly(A)-addition signal. In addition, the final Level III vector contains a Kan^R gene, whereas the monomers and multimers I-III hold resistance to ampicillin.

The two key requirements for finding a target site for TALENs are: a) binding site of each TALEN monomer should be preceded by a thymidine (T) and b) the distance of the binding site of each TALEN monomer, also known as the spacer, should be ~ 10–15 bases. According to the above requirements and the sequence context of the $HBB^{IVSI\text{-}110\ (G>A)}$ mutation, three binding sites for TALEN monomer two left, L1 and L2, and a common right, R1 were identified (Figure 47). The binding site of the R1 monomer is immediately downstream of the $HBB^{IVSI\text{-}110(G>A)}$ mutation and therefore the DSB produced by the HBB TAL L1/R1 or L2/R1 pairs will occur upstream of the mutation, approximately in

the middle of the two spacers, 10 bp and 13 bp, respectively. Based on the size and direction of the expected INDEL produced from the induced NHEJ DNA repair pathway following DSB formation, the IVSI-110(G>A) abnormal SA site should be disrupted, forcing the spliceosome to use the normal SA site as a result to achieve functional correction of the mutation. The binding sequence of the right TALEN monomer contains two repeats of GGG. The widely used monomers for binding G are those with the RVD NN. However, NN can also bind with less affinity to adenosine (A). The assembly kit offered an extra RVD, NK, which specifically binds to G but with less strength that the NN [184]. Therefore, we produced three modified versions of the R1 monomer (R2, R3 and R4) in which 2, 4 and 6 NN modules were replaced by NK (Figure 48).

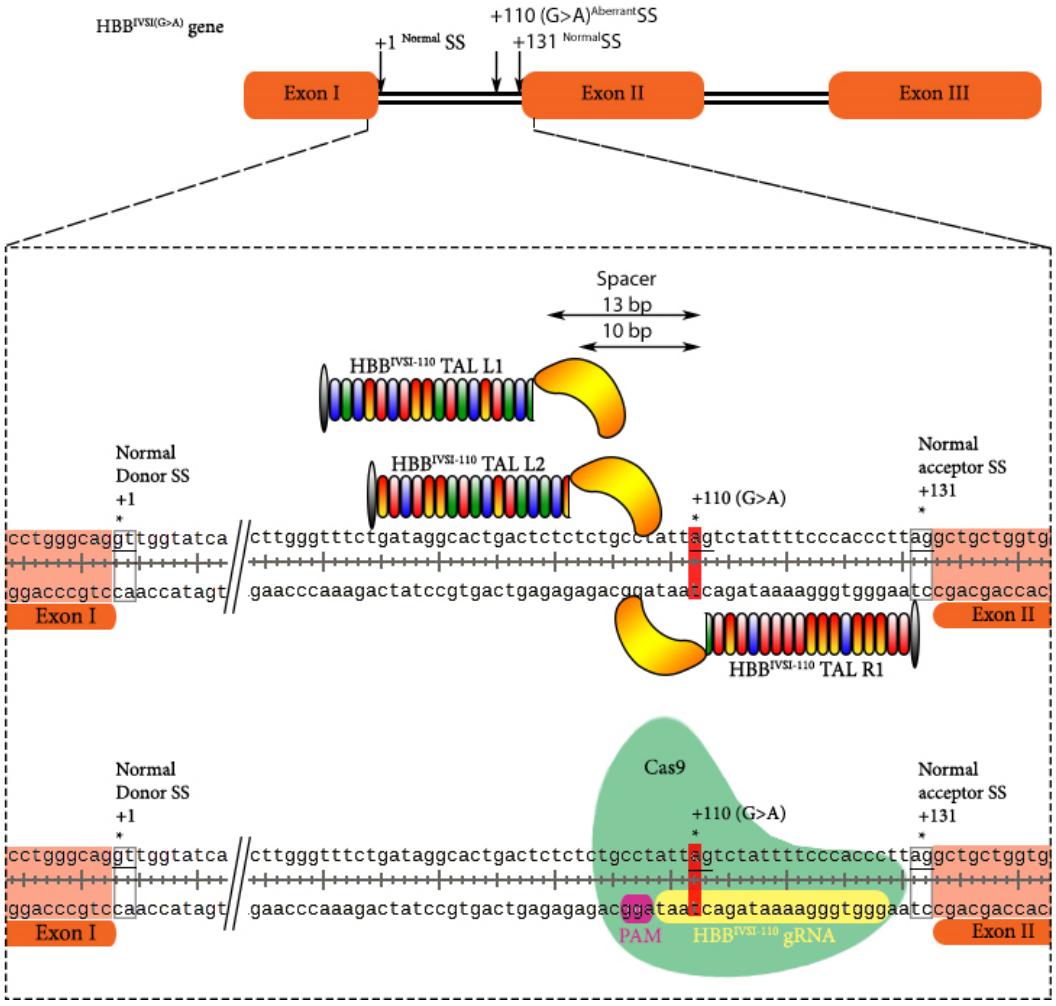


Figure 47: Schematic representation of the target sequences of HBB specific TALEN pairs R1/L1 and R1/L2 and the *HBB*$^{IVSI-110}$ specific RGEN on human *HBB*$^{IVSI-110}$ (G>A) gene.

In red highlight the *HBB*$^{IVSI-110}$(G>A) mutation is designated. HBB TALEN pairs R1/L1 and R1/L2 with the different spacer size, 13 bp and 10 bp, respectively. Green rounded shape: *Streptococcus pyogenes* Cas9. Yellow highlighted sequence: the 20 nt guide sequence of RGEN. Purple highlighted sequence: the NGG - PAM sequence. Rectangles indicate the normal donor and acceptor splice sites (SS).

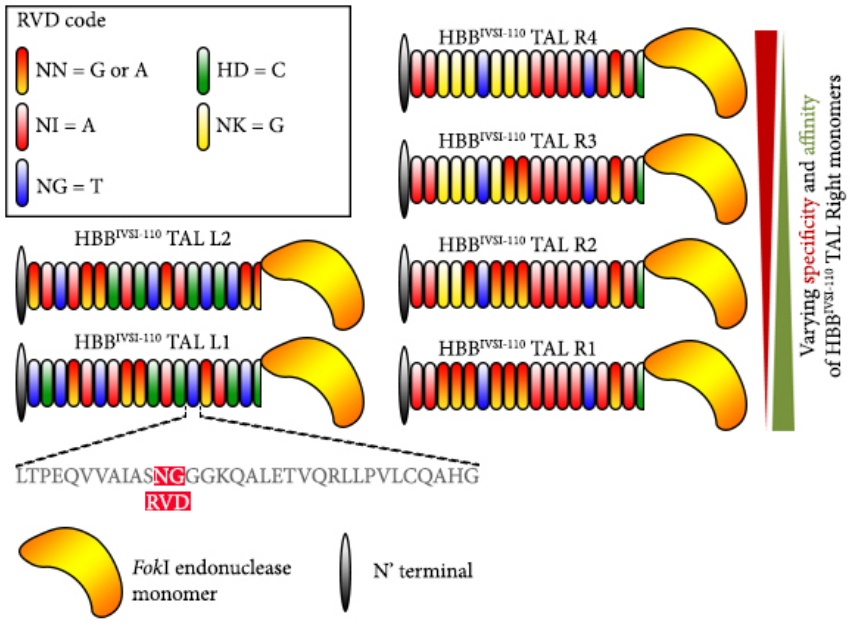


Figure 48: Schematic illustration of all of the produced *HBB* specific TALEN monomers.
Left 1 (L1), Left 2 (L2) and Right modified versions (1–4) monomers. Right monomers were modified by replacing the two repeats of three NN (G or A) modules (red to yellow gradient) with the more specific but less efficient NK (G) module (yellow). Substitution of NN to NK modules in the modified versions of Right TALEN monomer is: R1: 0 out of 6, R2: 2 out of 6, R3: 4 out of 6 and R4: 6 out of 6. Substitutions of NN to NK increase specificity and decrease the affinity of the Right TALEN monomer to their target sequence. Each TALEN monomer consists of 17.5 TALE repeats. Conserved amino acid sequence per TALE repeat in which the RVD (12th and 13th amino acid) is highlighted in RED. The yellow freeform shape is the *Fok*I endonuclease cleavage domain and the grey ellipse the N' TALE terminal. C' TALE terminal is not shown.

3.3.3.3 *In silico prediction of off-targeting potential of designer nucleases*

One of the major concerns with genome editing is the off-targeting potential of the designer nucleases, especially in the case of RGEN, which acts as a monomer [179]. One of the great gene candidates for off-targeting of *HBB*-specific designer nucleases is the highly homologous *HBD* gene. In a study which used an RGEN specific for the region close to sickle-cell disease mutation at codon 6 of *HBB* that the RGEN had off-target events, especially on the highly homologous *HBD* gene, which even caused gross chromosomal deletions [248]. In addition, in another investigation which identified specific RGEN target sequences on *HBB* and predicted their off-target effects, the authors suggested that the best region for genome editing of *HBB* are the introns, which have less conserved homology sequences than exons relative to *HBD* [249]. In the present study, the target sequence of both of the designer nucleases, RGEN and TALENs, is in *HBB* intron 1, which reduced the off-targeting potential on *HBD* and elsewhere. Furthermore, different *in silico* tools have been developed that allows the prediction of the off-targeting potential of designer nucleases [190].

3.3.3.3.1 Off-targeting potential of RGEN

In general, RGEN is not expected to be active at *HBD* since alignment of $HBB^{IVSI\text{-}110(G>A)}$ with *HBD* sequences indicates that there are 6 mismatches between the target sequence of the *HBB* RGEN and *HBD*, with five of them within the crucial seed sequence of the gRNA, which is less tolerant to mismatches. In addition, alignment skips two base pairs on the *HBD* sequence and therefore no PAM sequence on *HBD* (Figure 49). Prediction of $HBB^{IVSI\text{-}110}$ specific RGEN off-target potential using the MIT CRISPR design web tool (http://crispr.mit.edu/), estimated 264 off-target sites in the human genome of which 15 are in genes. As expected the highest score (38) is against normal *HBB* with 1 mismatch (the $HBB^{IVSI\text{-}110\ (G>A)}$), whereas the rest of the genes scored ≤1 and none of them in HBD (Table 20).

3.3.3.3.2 Off-targeting potential of TALENs

Analysis of the *HBB*-targeting TALEN pairs off-target potential using the PROGNOS web tool (http://bao.rice.edu/Research/BioinformaticTools/prognos.html), set with tolerance of up to 6 mismatches per nuclease half site (default) and spacer length (8–30bp), 3260 and 4303 off-target sites were detected in the human genome of which 65 and 72 are in exons for the HBB TALEN pair R1-4/L1 and R1-4/L2, respectively. As expected the highest score for both pairs was against normal *HBB* with a homology score of 98 and TALEN score 100 and a high homology score, 58, for *HBD* as a potential off-target site with TALEN scores of 82.35, 81.34, 78.44 and 76.83 for combinations with R1, R2, R3 and R4, respectively, regardless of which Left monomer was used. Note that the reduction of the TALEN scores against HBD follows the TALEN pairs more specific right monomers, R1 to R4 (Appendix section 6.1, Table 12-Table 19).

One parameter that might play a crucial role in reducing the off-targeting on the *HBD* gene beside the use of modified versions of the common right monomer is the difference of spacer length when L1 or L2 monomers are used. According to PROGNOS analysis and the alignment of $HBB^{IVSI\text{-}110}$ and *HBD* gene, the *HBD* sequence has two bases pairs less in the target site; which subsequently reduced the length of the R-/L1 and R-/L2 spacers to 10bp and 8 bp on the *HBD*, respectively. The length of the spacer contributes to the efficiency of TALENs, in which its been reported as optimal length to be the 12 bp (12–15 bp), whereas when the spacer lower than 10 bp is used the targeted disruption efficiency is impaired [228] (Figure 49). In fact, in initial PROGNOS analysis of the R-/L2 pairs, using the default spacer length settings (10–30 bp), the *HBD* gene was not included in the list of potential off-target sites.

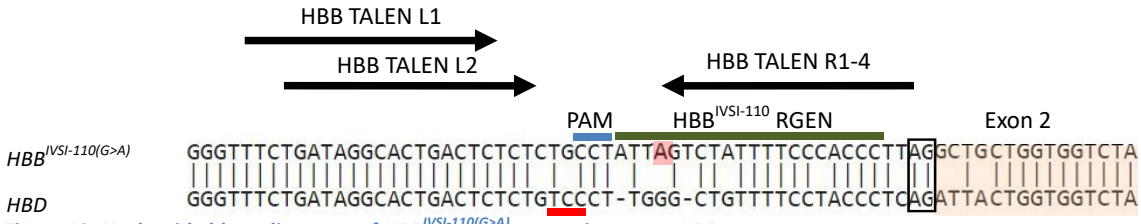


Figure 49: Nucleotide blast alignment of *HBB*$^{IVSI-110(G>A)}$ target site verses *HBD*.
Binding sites on *HBB*$^{IVSI-110(G>A)}$ are indicated as arrows for *HBB* TALEN monomers, green line for *HBB*$^{IVSI-110}$ RGEN (PAM sequence in blue). The *HBB*$^{IVSI-110(G>A)}$ mutation is highlighted in red. Exon 2 sequences are highlighted in orange and the normal splice acceptor sites are in rectangles for both aligned genes. Red line indicates the expected site of the PAM sequence on *HBD*, which is not functional since it not NGG.

3.3.3.4 *Characterisation of TALEN constructs at a protein level*

TALENs are relatively large proteins and therefore besides the typical confirmation at the plasmid DNA level by restriction enzyme test digests and Sanger sequencing (data not shown), the expression of the expected protein size (~110 KDa) was confirmed by immunoblots on protein extracts (30μg) of transfected HEK293T cells with single TALEN monomer constructs (Figure 50). TALENs are expressed from the pVAX vector with HA-TAQ fused at the N-terminal of the proteins, which is used for the detection of the each TALEN on immunoblots using an anti-HA antibody. All the newly developed *HBB*$^{IVSI-110}$ TALENs were transiently transfected to HEK293T cells in parallel with already characterised functional constructs (*CCR5* TAL Right and Left) to serve as positive controls [227]. Protein extracts from untransfected or cells transfected with pUC118 plasmid were used as negative controls. All *HBB*$^{IVSI-110}$ TALEN constructs expressed similar sized TALEN proteins to the CCR5 TALEN monomer controls; whereas no band was detected in the negative control samples (Figure 50).

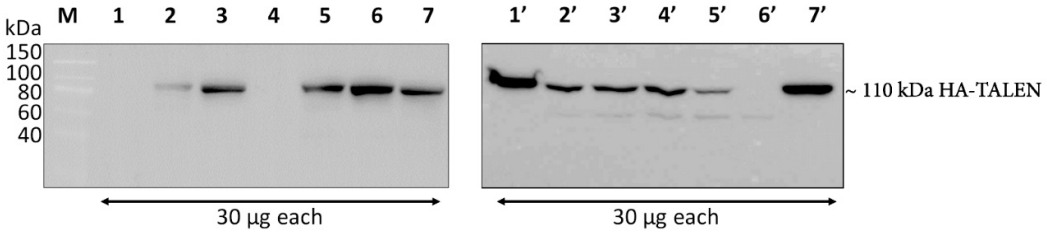


Figure 50: Immunoblot confirmation of TALEN monomers.
Evaluation of the correct protein size (~110 kDa) of the novel *HBB*$^{IVSI-110}$ TALEN monomers, L1 (lane 5), L2 (lane 6 and 7') and R1–4 (lane 7, 2', 3' and 4', respectively) from protein extracts (30 μg) from HEK293T cells transiently transfected with each TALEN construct separately. HEK293T cells were transfected in parallel with the already characterised functional constructs, *CCR5* TAL Right (lane 2, 3 and 5') and Left (lane 1'), as positive controls. Protein extracts from untransfected (lane 1 and 6') or cells transfected with pUC118 plasmid (lane 4) were used as negative controls. TALEN proteins were detected with an anti-HA antibody.

3.3.4 *Characterisation of targeted disruption efficiency of HBB$^{IVSI-110}$ - specific designer nucleases in HEK293T cells*

Designer nuclease targeted disruption efficiency was assessed by transient transfection of HEK293T cells using polyethylenimine. Targeted disruption efficiency of the designer nucleases was assessed with two methods; T7E1 assay and flow cytometry using a HBB$^{IVSI-110}$-GFP reporter construct.

3.3.4.1 *Targeted disruption efficiency of designer nucleases on HEK293T genomic DNA*

The T7E1 assay is widely used as a method to quantify the percentage of DSB-induced INDELs in the gDNA of the edited cell population. Briefly, in edited bulk cell populations the region encompassing the *HBB$^{IVSI-110\ (G>A)}$* target mutation is amplified by PCR, and therefore slightly different size of PCR products are produced, due to the presence of INDELs. When the chimeric PCR population is denatured and reannealed, a fraction of the PCR product population anneals with different size products, which leads to the formation of single strand loops at the site were the INDEL and where the DSB originally occurred. With the addition of T7 endonuclease 1, these loops are cleaved and the single product is cleaved into two smaller sized fragments, which are then resolved by agarose gel electrophoresis and quantified based on their band intensity. The ratio of cleaved versus uncleaved bands indicates the percentage of INDELs in the population and subsequently the percentage of edited cells, which can also be interpreted as targeted disruption efficiency of each designer nuclease.

Initially, we tested the efficiency of the *HBB$^{IVSI-110}$* - specific designer nucleases, HBB TAL R1/L1 and R1/L2 pairs, and HBB$^{IVSI-110}$ RGEN, in HEK293T cells. We used the *CCR5* - specific TALEN and RGEN as positive controls for the methods and pUC118 and untransfected cells as nuclease-free negative controls. Cells were co-transfected with equal amounts of a pRK5.mCherry construct (50 ng) for the assessment of transfection efficiency. Transfection efficiency was similar in all samples (~55-62%) and cell death remained below 12% (measured by Luna Automated cell counter with trypan blue staining) (Figure 52). Targeted disruption of the *HBB* locus reached up to 24.5%, 19% and 10.4% for the HBB$^{IVSI-110}$ TAL R1/L1, R1/L2 and HBB$^{IVSI-110}$ RGEN, respectively.

Off-targeting of the *HBB$^{IVSI-110}$* specific-designer nucleases on the highly homologous *HBD* was close to background with cleaved bands of 4.8%, 2.19% and 0.9% for HBB TAL R1/L1, R1/L2 and RGEN (Figure 51 & Figure 52). These results are in agreement with the *in silico* predicted off-targeting potential (section 3.3.3.3.1), where the highest possibility for off-targeting on *HBD* was with HBB TALEN pair R1/L1. Noteworthy is the fact that HEK293T cells have normal *HBB*, whereas the HBB$^{IVSI-110}$ RGEN target sequence includes the *HBB$^{IVSI-110(G>A)}$* mutation and this mismatch might impair its

activity. *CCR5* specific TALEN and RGEN negative controls showed ~10.4 % and 3.3 % targeted disruption efficiency on *CCR5* (Figure 51 & Figure 52).

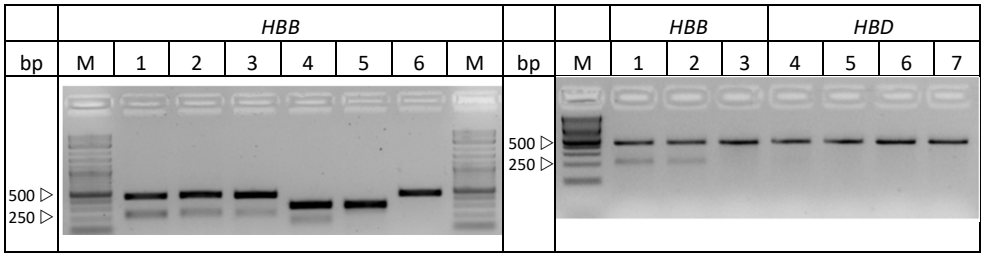


Figure 51: Targeted disruption efficiency of designer nuclease on HEK293 gDNA by T7E1 assay.
Left panel: agarose gel electrophoresis analysis of PCR products after treatment with T7 endonuclease 1. Lanes 1,2, 3 and 6 are *HBB* PCR products derived from transfected HEK293T cells with HBB TAL R1/L1, R1/L2, RGEN and pUC118, whereas 4 and 5 are *CCR5* PCR products from cells transfected with CCR5 TALEN and RGEN, respectively. Right panel shows targeted disruption efficiency of HBB TALEN R1/L1 (lane 1) and R1/L2 (lane 2) on the *HBB* locus relative to the negative control pUC118 transfected cells (lane 3). Lanes 4, 5, 6 and 7 show the parallel off-target potential of the both HBB TALEN pair R1/L1 and R1/L2 and RGEN, relative to the nuclease-free negative control, pUC118, on the highly homologous *HBD* respectively.

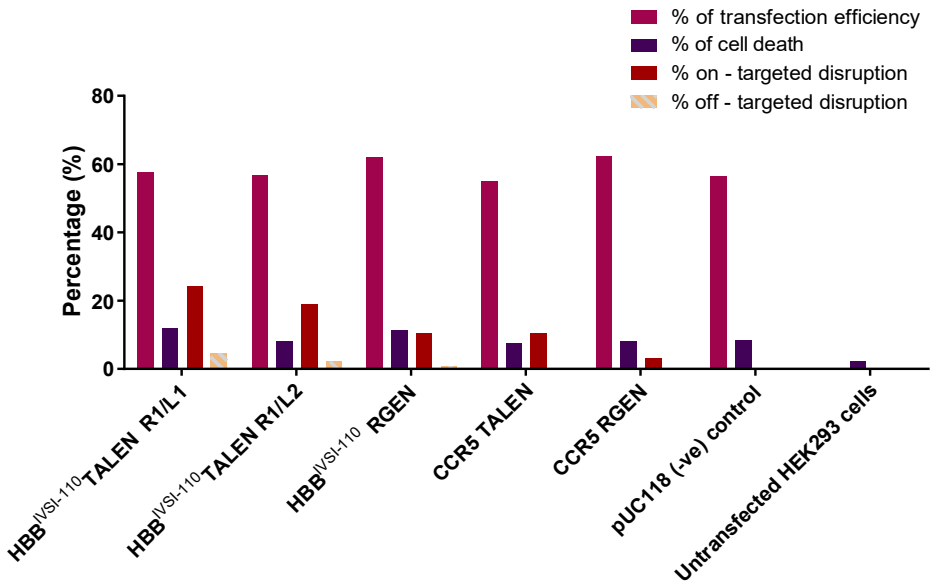


Figure 52: Genome editing on HEK293T cells.
Analysis of transient transfection efficiency (mCherry positive, pink bars) and cell death (trypan blue positive, purple bars) of HEK293T cells with designer nucleases at 48hr post-transfection. Red bars: on-target disruption efficiencies. Orange bars: off-target potential of *HBB* specific designer nucleases TALEN R1/L1 and R1/L2 and HBB$^{IVSI-110}$ RGEN on *HBB* and *HBD* loci, respectively. *CCR5* specific TALEN and RGEN are included as positive controls of the method and accordingly on-targeted disruption efficiencies on *CCR5* are shown in red bars. pUC118 is used as nuclease-free negative control.

3.3.4.2 *Targeted-disruption efficiency of specific designer nucleases on an episomal HBB$^{IVSI-110}$-GFP reporter construct in HEK293T cells*

Although the T7E1 assay is reliable and widely used it is not very accurate since it is based on band quantification following resolution of products by agarose gel electrophoresis and thus characterised

by low sensitivity. As described above several modified versions of the common HBB$^{IVSI-110}$ TAL right monomer were designed, R1–4, with varying specificity and affinity to their binding target sequence. Therefore, a more sensitive method was needed in order to detect even minor differences in activity between the different combinations of R1–4 with L1 or L2 monomers. To this end an HBB$^{IVSI-110(G>A)}$-GFP reporter gene construct, in which we cloned the *HBB*$^{IVSI-110(G>A)}$ target region in frame immediately upstream of the eGFP cDNA was generated (Figure 53).The whole construct was driven by the U6 promoter. Disruption of the *HBB*$^{IVSI1-110}$ region by the designer nucleases should lead to the formation of INDELs in which statistically the 2/3 change the ORF of eGFP leading to a drop in the percentage of GFP positive cells, easily and accurately detected by flow cytometry. This reporter construct was transiently co-transfected into HEK293T cells with the *HBB*-specific designer nucleases (TALENs and RGEN), along with a mOrange transfection reporter construct. In addition, cells were transfected with GFP-specific designer nucleases (TALENs and RGEN) as positive controls, *CCR5*-specific designer nucleases (TALENs and RGEN) as nuclease negative controls and pUC118 construct and untransfected cells as nucleases-free negative controls. Cell death caused by the constructs or the transfection procedure was assessed by flow cytometry after staining with SYTOX Red (cell death stain).

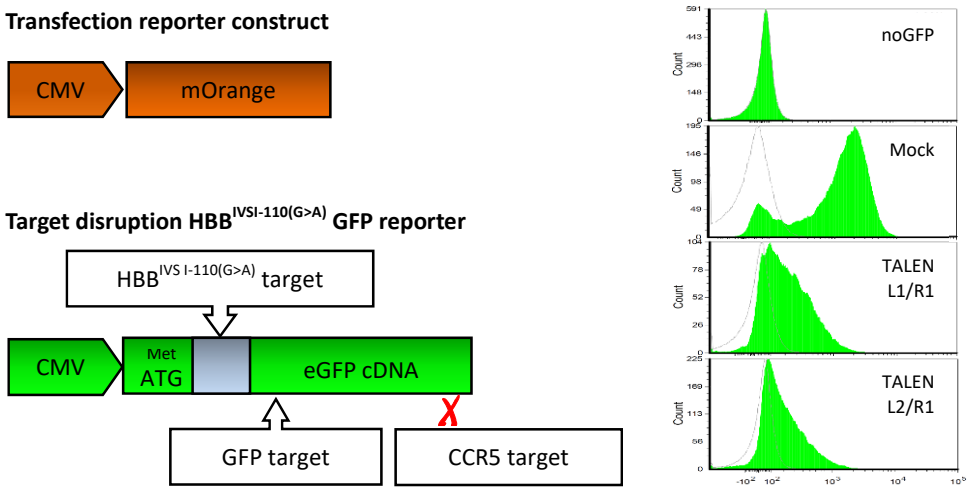


Figure 53: Schematic representations of the transfection reporter construct pCMV mOrange and targeted disruption HBB$^{IVSI-110(G>A)}$ GFP reporter.
The HBB$^{IVSI-110}$ GFP reporter was prepared by inserting the HBB$^{IVSI-110}$ target sequence between the start codon (ATG) and the eGFP cDNA. Therefore, GFP HBB$^{IVSI-110(G>A)}$ can be used as a target construct for HBB$^{IVSI-110}$-specific designer nucleases for the assessment of targeted disruption efficiencies, as reduction of the percentage of GFP positive cells or mean fluorescence intensities in the mOrange live (SYTOX Red-negative) population (see histograms). GFP and CCR5 specific designer nucleases can be used as positive and negative controls, respectively.

A total of three independent experiments were performed with HEK293T cells, with each treatment in duplicate. Transfection efficiencies were similar among samples, on average 43.6±6.67% and cell death remained below 10% (Figure 54). In order to avoid differences in the GFP positive cell

population due to differences in transfection efficiencies or cell death, we focused our analysis only on the mOrange-positive and live (SYTOX Red-negative) population. All *HBB*$^{IVSI-110}$ TALEN pair combinations showed a significant reduction in the percentage of GFP positive cells in the live mOrange (positive) HEK293T population. In particular, samples treated with the unmodified version of right monomer, *HBB*$^{IVSI-110}$ TAL R1 (no NN>NK substitution), paired with *HBB*$^{IVSI-110}$ HBB L1 and L2 monomers, the reduction reached up-to 79% (*p* value = 0.0001) and 86% (*p* value = 0.0001) relative to the HBB$^{IVSI-110}$-GFP reporter nuclease-free negative control, respectively, exceeding even the activity of the GFP-specific designer nucleases, 40% TALEN and 43% RGEN (*p* value = 0.0001). Pairs with the *HBB*$^{IVSI-110}$ TAL R2 (substitution of 2 out of 6 NN>NK modules) showed lower activity than TALEN combinations with R1 monomer, by ~40% (*p* value = 0.0001) and ~55% (*p* value = 0.0001) when paired with L1 and L2, respectively. Combinations of both left TAL monomer with R3 (substitution of 4 out of 6 NN>NK modules) and R4 (substitution of 6 out of 6 NN>NK modules) monomers showed even less activity, 16% (*p* value = 0.0055), 17% (*p* value = 0.0022) for L1/R3 and L1/R4 and 50%, 55% (*p* value = 0.0001) for L2/R3 and L2/R4, but with no significant difference between them. In addition, all *HBB*$^{IVSI-110}$ TAL pair combinations with L2 monomer (10 bp spacer) showed higher targeted disruption efficiency than those with L1 (13bp spacer). This might be due to higher binding efficiency of the TAL L2 monomer than L1 and/or because of the shorter spacer, which could increase the frequency of *Fok*I dimerization. In the case of *HBB*$^{IVSI-110}$ RGEN the percentage of GFP positive cells in the live mOrange positive population dropped by 25% (*p* value = 0.0001) relative to the nuclease-free negative control. The percentage of GFP^{+} cells in both *CCR5* specific designer nucleases, TALENs and RGEN, as expected, remained close to the background negative control levels (Figure 55).

As described above, targeted disruption activity was also quantified as the reduction of GFP mean fluorescence intensity (MFI) in the live mOrange positive cell population relative to the nuclease-free negative control. Comparison of the MFI instead GFP^{+} cell numbers increased the sensitivity of the method since the level of GFP expression was also taken into account rather than just distinguishing the GFP positive and negative cells. Results from the MFI comparison showed a similar pattern of disruption to that observed based on the comparison of percentage GFP positive cells but at an even higher level of efficiency. In addition, comparison of HBB$^{IVSI-110}$ TALEN pairs, showed that substitution of 2 out of 6 NN>NK module (R2 monomer) significantly decreased targeted disruption efficiency when paired with L1 or L2 monomers to the ½ relative to those paired with unmodified R1 monomer (L1/R1 (16.8%) > L1/R2 (35%); L2/R1 (15%) > L2/R2 (26.8%)(*p* value = 0.0001)). Moreover, substitution of NN>NK modules on the HBB$^{IVSI-110}$ Right monomer (substitutions of 4 and 6 out of 6 in R3 and R4 monomers, respectively) further decreased the activity in pairs with L1 monomer by 2/3

(*p* value = 0.0001) relative to pairs with R1, whereas in combination with L2 monomer no further reduction was observed. No significant changes in the activity of TALEN pairs were seen with or without the last two NK modules (Figure 55). *HBB*$^{IVSI-110}$ RGEN showed relatively low activity, in which 74% and 48% (*p* value = 0.0001) reduction was observed in the percentage of GFP positive cells and GFP MFI in the live mOrange positive HEK293T cell population relative to the nuclease-free negative control, respectively (Figure 55). GFP specific TALEN and RGEN positive controls showed a significant reduction in the percentage of both GFP^{+} cells (~59%; *p* value = 0.0001) and MFI (~35%; *p* value = 0.0001), in live mOrange-positive cells relative to the nuclease-free negative control. In contrast, the *CCR5* specific TALEN and RGEN negative controls gave GFP and MFI levels close to nuclease-free negative control levels, except in the case of the GFP MFI levels in which the CCR5 TALEN exhibited a modest reduction by 13.5 % (*p* value = 0.0001) (Figure 55).

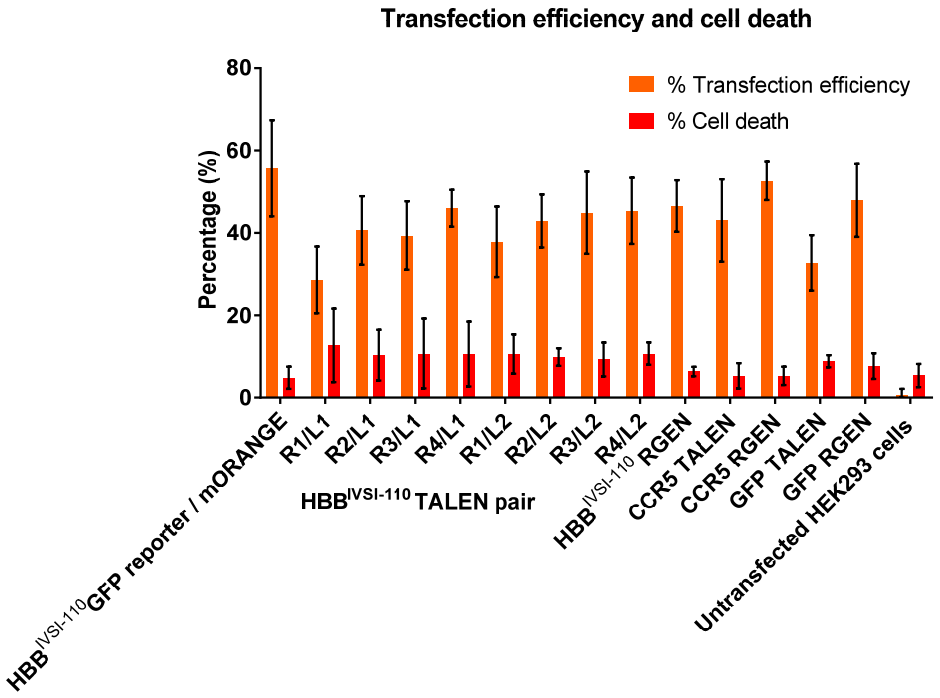


Figure 54: Transfection efficiencies measured as mOrange positive cells (orange bars) and cell death as SYTOX Red-positive cells (red bars) measured via flow cytometry.
Experiments were conducted in the triplicate by transient transfection of HEK293T cell for the assessment of targeted disruption of the HBB$^{IVSI-110}$-GFP reporter gene construct. Cells were transfected with all HBB TALEN combination (L1/R1–4 and L2/R1–4), the HBB$^{IVSI-110}$ RGEN, CCR5 TALEN and RGEN, and GFP TALEN or RGEN, with CCR5- and GFP- specific nucleases to serve the designer nuclease negative- and positive- controls of the method respectively. All cells were co-transfected with equal amounts of mOrange construct for the assessment of transfection efficiency. HBB$^{IVSI-110}$ GFP reporter/mOrange sample serve the nuclease-free negative control.

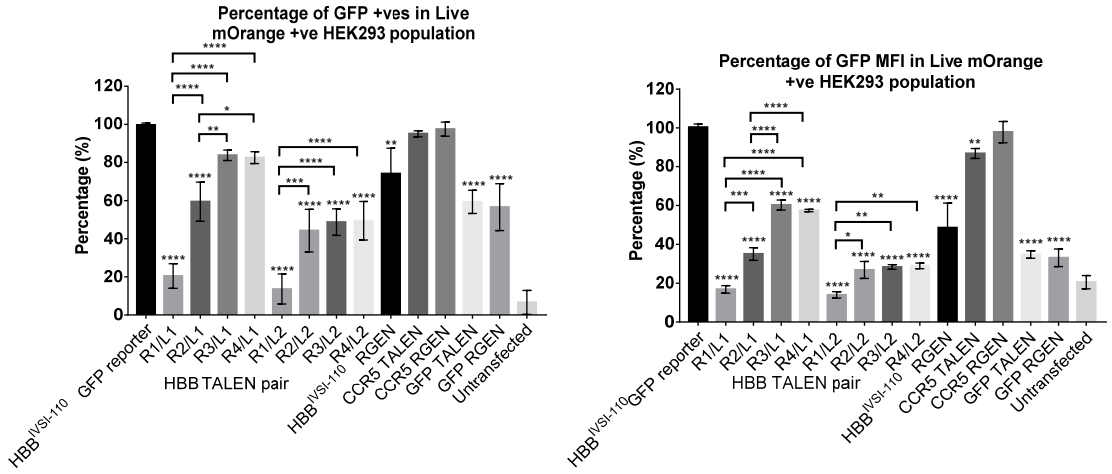


Figure 55: Assessment of targeted disruption efficiency of designer nucleases on the episomal HBB$^{IVSI-110}$ GFP reporter via flow cytometry.

Experiments were conducted in triplicate by transient transfection of HEK293T cell for the assessment of targeted disruption of the HBB$^{IVSI-110}$-GFP reporter gene construct. Cells were transfected with all HBB TALEN combinations (L1/R1–4 and L2/R1–4), the HBB$^{IVSI-110}$ RGEN, CCR5 TALEN and RGEN, and GFP TALEN or RGEN, with CCR5- and GFP- specific nucleases to serve the designer nuclease negative- and positive- controls of the method, respectively. Right panel: targeted disruption efficiency of the HBB$^{IVSI-110}$-GFP reporter gene construct quantified as a reduction of the percentage of GFP positive cells in the live (SYTOX Red negative) and mOrange positive transiently transfected HEK293T cell population compared to the nuclease free negative control (HBB$^{IVSI-110}$GFP reporter). Data obtained from duplicates in three independent experiments. Left panel: targeted disruption efficiency quantified as a reduction of the percentage of GFP MFI in live (SYTOX Red negative) and mOrange positive HEK293T cell population compared to the nuclease free negative control (HBB$^{IVSI-110}$GFP reporter). Data obtained from duplicates in three independent experiments. Statistical differences in the percentage of GFP positive cells and GFP MFI for all samples was tested compared to the HBB$^{IVSI-110}$ GFP reporter/mOrange negative control. Moreover statistical analysis was tested between HBB TALEN combinations with L1/R1–4 and L2/R1–4, separately. Statistical analysis was performed by one-way ANOVA.

* $p < 0.05$, ** $p < 0.01$*** $p < 0.001$, **** $p < 0.0001$.

Percentage of GFP positive cells	
Statistical analysis:	
HBB$^{IVSI-110}$ GFP reporter compared to:	
All	**** – p value 0.0001
HBB TAL R3/L1	** – p value 0.0055
HBB TAL R4/L1	**– p value 0.0022
HBB TAL R1/L1 compared to:	
HBB TAL R2, R3, R4 / L1	****– p value <0.0001
HBB TAL R2/L1 compared to:	
HBB TAL R3, R4 / L1	**** – p value <0.0001
HBB TAL R1/L2 compared to:	
HBB TAL R2, R3, R4 / L2	****– p value <0.0001

Percentage of GFP MFI	
Statistical analysis:	
HBB$^{IVSI-110}$ GFP reporter compared to:	
All	**** – p value 0.0001
HBB TAL R1/L1 compared to:	
HBB TAL R2, R3, R4/L1	****– p value <0.0001
HBB TAL R2/L1 compared to	
HBB TAL R3, R4 / L1	**** – p value <0.0001
HBB TAL R1/L2 compared to	
HBB TAL R2, R3, R4 / L2	**** – p value <0.0001

3.3.5 *Assessment of functional correction potential of designer nucleases in an HBB$^{IVSI-110}$ transgenic MEL cell line*

Subsequent to the assessment of targeted disruption efficiency of both HBB$^{IVSI-110}$-specific TALEN pairs and RGEN in HEK293T cells (sections 3.3.4.1 and 3.3.4.2), functional correction potential of the mutation was initially assessed using MEL cells containing stably integrated copies of the MA821 HBB$^{IVSI-110}$ LV transgene, before moving to patient-derived CD34^{+} cells. The transgenic MEL MA821 HBB$^{IVSI-110}$ cell pool was produced by transduction of MEL cells with a modified version of the MA821 LV, in which we inserted the *HBB$^{IVSI-110\ (G>A)}$* mutation in the *HBB* transgene via a site directed mutagenesis kit.

The bulk transduced MEL cell population with an average VCN of 1.9 was selected to perform genome editing experiments. Characterisation of the expression profile of the MA821$^{IVSI-110\ (G>A)}$ cell pool was assessed at an mRNA level via multiplex RT-qPCR in which the two HBB mRNA variants were quantified showing 45% of correctly spliced HBB mRNA with the remainder being aberrantly spliced (Figure 25). Analysis at the protein level showed, as expected a low degree of functional HBB chains at ~5% compared to the normal counterpart MA821 (VCN ~2) (Figure 26). Since MEL cells grow in suspension, electroporation of plasmid DNA was used as a transfection method for introducing the designer nucleases into these cells, with a series of optimisation experiments being performed for the adjustment of the amount of plasmid DNA, number of cells and electroporation conditions (voltage and capacitance), before proceeding to the genome editing experiments (data not shown). Briefly, the transfection procedure consisted of electroporation of ~5 x 10^6 cells with 60 µg total plasmid DNA of which 1/3 is a CMV-GFP transfection efficiency reporter construct and the remaining 2/3 is the designer nuclease construct or pUC118 nuclease-free negative control, in 0.4cm cuvettes and final volume of 500 µL. The optimal electroporation settings in terms of transfection efficiency and cell viability were: a) voltage at 400 kV, b) capacitance at 1050 µF and c) infinite resistance. Only the best combinations of *HBB$^{IVSI-110}$* specific TALEN pairs, R1/L1 and R1/L2, and the *HBB$^{IVSI-110}$* RGEN were evaluated in this system, even though the latter showed lower levels of disruption. Transfection efficiency and viability was assessed by flow cytometry after 72hr of hypothermic culture condition (30 °C) post-electroporation. Genomic DNA for targeted disruption assessment was collected after a 3–4 day recovery period under standard culture conditions (37 °C).

3.3.5.1 *1st experiment on MEL MA821$^{IVSI-110}$ transgenic cells pool*

As a pilot experiment, two sequential electroporation transfections of the MEL MA821$^{IVSI-110}$ transgenic cell pool was performed with the same designer nucleases in order to increase the fraction of genome edited populations.

3.3.5.1.1 Assessment of transfection and targeted disruption efficiencies

The first round of electroporation showed a similar percentage of live (SYTOX Red-positive), GFP (positive) cells, 43.6%, 44.4%, 43.3% and 48.8% and relatively high cell death (SYTOX Red-positive) 58.5 % 52,2% , 51.8% and 36.8% for electroporated samples with *HBB$^{IVSI-110}$* TALEN R1/L1 and R1/L2, *HBB$^{IVSI-110}$* RGEN and the pUC118 nuclease-free negative control, respectively (Figure 57). This degree of cell death is typical for electroporation.

A T7E1 assay showed 15.6% and 12.2% targeted disruption in cells edited with the two *HBB$^{IVSI-110}$* TALEN pairs R1/L1 and R1/L2 respectively. Contrastingly, the *HBB$^{IVSI-110}$* RGEN construct did not show any activity (Figure 56 & Figure 57). When cells recovered from the first round of electroporation, a sample of the transfected MEL cell populations were subjected to a second round of electroporation with the same constructs. Following the two-rounds of electroporation *HBB$^{IVSI-110}$* TALEN R1/L1 and R1/L2, *HBB$^{IVSI-110}$* RGEN and the pUC118 nuclease-free negative control cells showed a very low level of cell death namely 0.05%, 0%, 2.59% and 0% (SYTOX Red-positive cells) and transfection efficiency (GFP$^+$ cells) 60%, 40%, 32.4% and 38.5%, respectively (Figure 56 & Figure 57). Comparison of the targeted disruption in two-round transfected edited cells relative to the single round indicated a noticeable increase only in the case of *HBB$^{IVSI-110}$* TAL R1/L1, which reached up to a 19.5% genome edited population, whereas the rest remained approximately the same at 11.8% and 3.2% for *HBB$^{IVSI-110}$* TAL R1/L2 and RGEN, respectively (Figure 57).

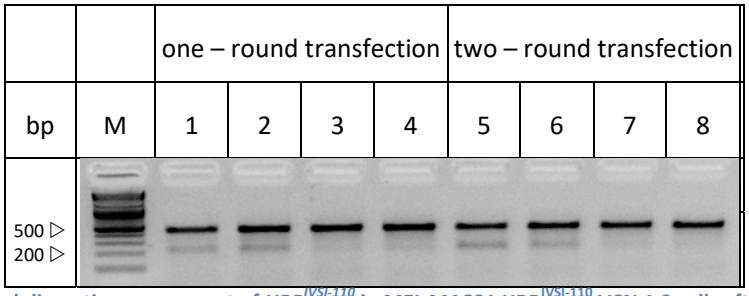


Figure 56: Targeted disruption assessment of *HBB$^{IVSI-110}$* in MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cells after the one round and two rounds of electroporation with designer nucleases (1st Exp).
Following transfection of cells with the designer nucleases after one or two round of electroporation, total DNA was subjected to a T7E1 assay and products resolved by agarose gel electrophoresis. Lanes 1, 5: HBB TALEN R1/L1, 2, 6 HBB TALEN R1/L2, 3, 7 HBB$^{IVSI-110}$ RGEN and 4, 5 pUC118 nuclease free negative controls. M: 2log DNA ladder. Uncleaved band 500 bp, cleaved band 200 bp. Bands were quantified using ImageJ software.

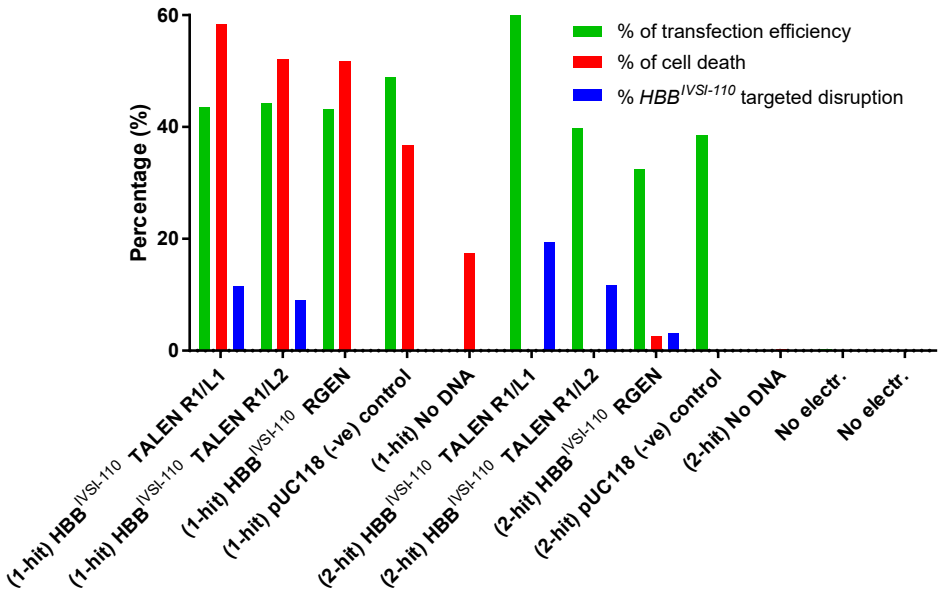


Figure 57: Genome editing of MEL HBB$^{IVSI-110}$ VCN1.9 cells with one (1-hit) and two (2-hit) rounds of transfection by electroporation with designer nucleases (1st Exp).
10x10^{6} cells were electroporated with 60 μg of plasmid DNA (40 μg designer nuclease/pUC118, 20 μg CMV GFP reporter). Electroporated cells were cultured for 72 hours at hypothermic conditions and then harvested for assessment of transfection efficiency (% of GFP positive cells) via flow cytometry, percentage of cell death (trypan blue positive cells) via microscopy, and percentage of genome edited cells in each population measured as targeted disruption by T7E1 assays (blue bars).Electroporated cells without DNA (No DNA) and no electroporated cells (No electr.) were used as GFP negative controls for the assessment of transfection efficiency and pUC118 nuclease-free negative controls for the assessment of targeted disruption.

3.3.5.1.2 Restoration of correct pre-mRNA splicing

Single and double transfection genome-edited MEL cell populations were induced to undergo terminal erythroid differentiation, in parallel with the pUC118 nuclease-free non-edited MEL MA821$^{IVSI-110}$ cells as a negative control. Cells were harvested on days 3 and 6 post-differentiation and total RNA extracted. Assessment of the restoration of splicing was achieved via an optimised multiplex RT-qPCR based on plasmid SC and two HBB variant specific probes as described (section 3.1.2.1). Briefly, using a plasmid SC of the pCR2.1 HBB cDNA^{Normal} cDNA^{Aberrant} plasmids of known quantities were used to accurately quantify specifically the number of molecules of both HBB mRNA variants (correct and aberrant) from which an estimate of the contribution of each variant as a percentage in the total population was determined. In addition, the efficiency of restoration of correct splicing in edited cells is assessed as the percentage change of each variant relative to the non-edited negative control, pUC118. On day-3 of differentiation the pUC118 nuclease-free negative controls had a ratio of 1:1 of correct/aberrant (c/a) mRNA in the total population whereas, on day-6 the percentage of aberrantly spliced mRNA increased to ~64 %, ratio 0.57 (Figure 58, left panel). Cells edited with TALEN pairs changed the ratio of the negative control in favour of the correctly spliced HBB mRNA in both single-round and two-round electroporation. In contrast, treatment with

the RGEN construct did not show any evidence of restoration of correct splicing, except from the two-round transfection sample on day-6 of differentiation, which remained at 49% (ratio c/a: 1), whereas the counterpart negative control dropped to 36.5% (ratio c/a: 0.96) (Figure 58**,** left panel). Data from the percentage change of the two variant mRNAs relative to the pUC118 negative control showed an increase of the correctly spliced mRNA ranging from 15% to 55% in samples edited with both TALEN pairs (Figure 58, right panel). Furthermore, a comparison of samples electroporated once or twice in an attempt to increase the percentage of genome editing, showed a considerable increase of restoration of correct splicing only in the case of TAL R1/L1 pair. These data are in agreement with the percentage of targeted disruption (Figure 56 & Figure 57) in which two-rounds of transfection with TAL R1/L1 increased the percentage of genome edited cells from 12% to 20%, whereas all other constructs remained approximately the same. The one-fold increase of genome-edited population is perfectly represented with the one-fold increase of the correctly spliced mRNA. In addition, comparison of data from day-3 to day-6 post-differentiation in TALEN-genome-edited cells revealed a pattern in which the correctly spliced HBB mRNAs accumulated at higher rates relative to non-edited cells (Figure 58, right panel). This might be a result of the partial restoration of correct splicing via genome editing in combination to the higher stability of the correctly spliced HBB mRNA over the aberrant.

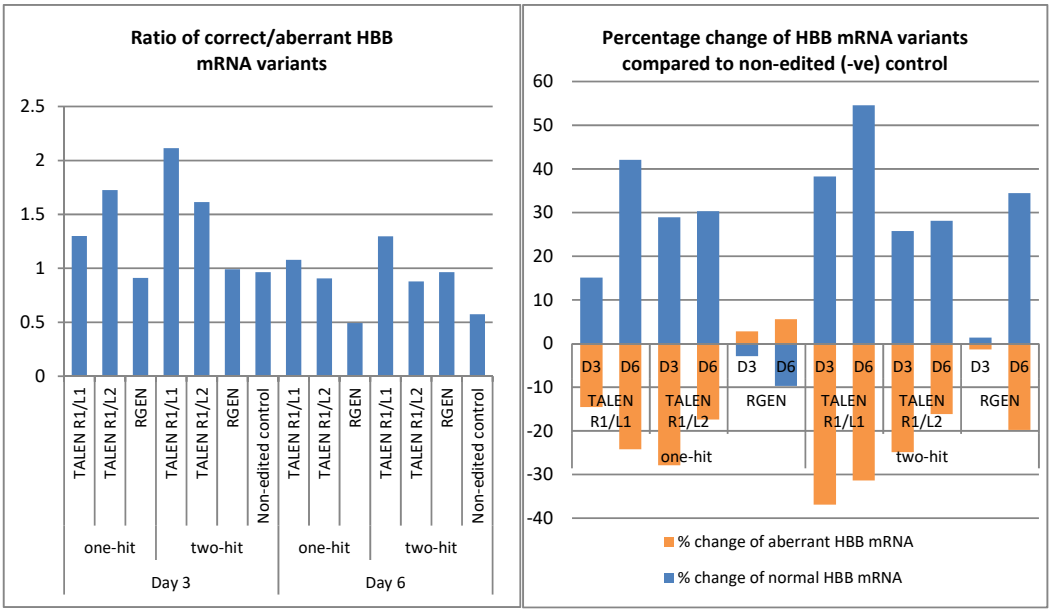


Figure 58: Assessment of the level of restoration of correct pre-mRNA splicing at day-3 (D 3) and day-6 (D 6) post-erythroid differentiation of genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cells (1st Exp).
MEL MA821 HBB$^{IVSI-110}$ cells were genome edited with designer nucleases (HBB TALEN R1/L1, R1/L2 and HBB$^{IVSI-110}$ RGEN) in one (1 –hit) and two (2-hit) rounds of electroporations. Genome edited cells were induced differentiated in induction medium (cRPMI suppl. 2% DMSO) and cells were collected for RNA analysis on day 3 and day 6.via multiplex RT-qPCR. Absolute quantities of correct and aberrant HBB mRNA variants were measured using a plasmid-based SC. Left panel: ratios of correct/aberrant HBB mRNA in the total HBB mRNA population. Right panel: percentage change of correctly (blue bars) and aberrantly (orange bars) spliced HBB mRNAs in genome edited samples relative to non- edited negative (–ve) control levels on day 3 (D3) and day 6 (D6) of differentiation.

3.3.5.1.3 Functional correction of HBB$^{IVSI-110}$ splicing at a protein level

Functional correction of genome edited MEL MA821 HBB$^{IVSI-110}$ cells was assessed at a protein level. Genome-edited cells were harvested for protein analysis by immunoblots on day-6 (day-9 where designated) of erythroid differentiation. MEL cells were used in this experiment as a negative control for HBB chain expression. Immunoblots from days 6 and 9 of differentiation showed more intense HBB chain bands in all samples treated with TALEN pairs R1/L1 and R1/L2, with greater amounts in the two-rounds of transfection samples, relative to the corresponding non-edited controls (Figure 59). The Hba chain bands gave similar intensities in samples from the same day and as expected continued to rise in parallel with differentiation (days 6 to 9) (Figure 59). Band densities of Actb (used to normalise protein loading), Hba chains (used to normalise for the level differentiation) and HBB chains (levels of correction) were quantified by ImageLab software. The level of correction was measured as fold-differences of differentiation-normalised (via Hba chains) of HBB chains relative to the non-edited negative control. Both globin chains were normalised for protein loading by reference to the Actb signal.

Analysis of band intensities on day-6 of differentiation, exhibit an increase in the level of HBB chains above the baseline of the non-edited controls in the two-round transfection edited samples with both TALEN pairs by ~0.6–0.7 fold and to a lower extent with RGEN (0.2 fold). The one-round transfection TALEN-edited samples, although HBB chains bands on immunoblots are more intense than the non-edited control, normalisation by reference to Hba and Actb levels gave a decreased correction value below the baseline of non-edited control. This may be due to a technical error following normalisation of Hba chains by reference to Actb levels which appears lower than expected since equal amounts were loaded on the blot for HBB quantification. This might be due to incomplete transference of the higher molecular weight proteins (such as Actb) in samples close to edge of the membrane during transfer. Contrastingly, differentiation levels measured on day-9 seem to be similar among samples and here correction applied at a protein level is more obvious in both 1- or 2-round transfected TALEN treated cells, with the latter producing 2.2 times more HBB chains relative to the non-edited negative control. Interestingly, a slight increase of differentiation-normalised HBB chains was detected in RGEN- treated samples but which is still close to background levels (~0.25 fold higher) (Figure 59, upper panel).

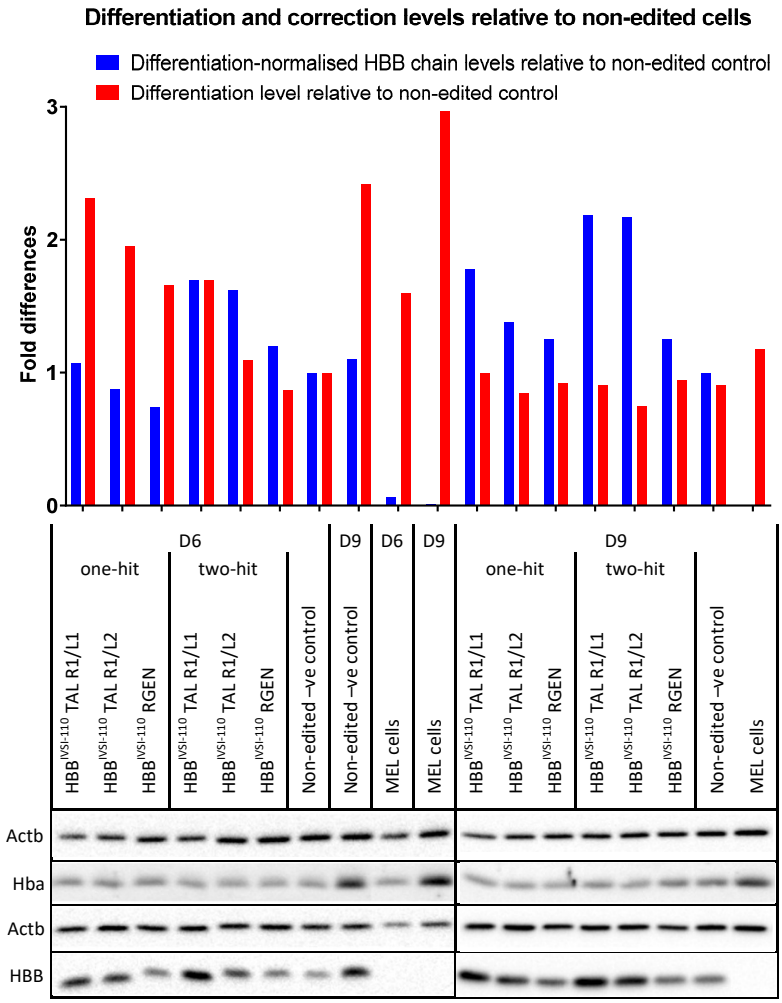


Figure 59: Immunoblots for quantification HBB, and Hba chains and Actb in protein extracts from day-6 (D 6) and day-9 (D 9) erythroid differentiated genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cells (1st Exp).
Lower panel: MEL MA821 HBB$^{IVSI-110}$ cells were genome edited with designer nucleases (HBB TALEN R1/L1, R1/L2 and HBB$^{IVSI-110}$ RGEN) in one (1 –hit) and two (2-hit) rounds of electroporations. Genome edited cells were induced differentiated in induction medium (cRPMI suppl. 2% DMSO) and cells were collected for protein expression analysis on day 6 and day 9 via western immunoblots analysis. Whole cell extracts from 0.5–1x10^{6} cells were probed with antibodies against HBB, Hba and Actb. Actb was used as a protein loading control. Non-edited controls were used as negative control for the assessment of level restoration of splicing of genome edited cells at the protein level, whereas MEL cells as negative control for the HBB chain expression. Band densities were measured using ImageLab. Upper panel: fold difference of differentiation-normalised HBB chain expression (via HBB / Hba chain ratio; blue bars) and differentiation levels (via Hba chains; red bars) compared to non-edited negative (–ve) controls.

3.3.5.2 *2nd experiment on MEL MA821$^{IVSI-110}$ transgenic cells*

Data from the 1st experiment with the MEL MA821$^{IVSI-110}$ cells provided a proof of principle for our hypothesis. At this point, TALENs exhibited encouraging results regarding restoration of correct splicing at a transcriptional level and consequent increased HBB chain production, even at a relatively low (~10%) targeted disruption percentage. However, a crucial negative control that was lacking from the 1st experiment (section 3.3.5.1) was cells treated with TALEN and RGEN nucleases against targets that are not associated in any way with erythropoiesis. Thus the aims of the 2nd experiment, besides replication of the observed results in the 1st experiment, was the reinforcement of these data by including *CCR5*-specific TALENs and RGEN treated cells as extra negative controls. Furthermore, in an attempt to increase the percentage of the genome-edited cell population, DNA end-processing enzymes such as the three prime repair exonucleases 2 (TREX2), were also included in the treatment protocol. TREX2 is a nuclease protein with a 3' to 5' exonuclease activity and is involved in the DNA repair pathway of DNA DSBs. It has been shown that co-transfection with designer nucleases such as TALEN, ZFN and especially homing endonucleases, TREX2 increases targeted disruption frequencies by up to 25-fold [250]. Upon a DSB being produced by designer nucleases such ZFN, TALENs and homing endonucleases, formation of compatible overhangs occurs, which can be precisely re-ligated without further processing leading to an accurate repair of the DSB. TREX2 can be used to increase the percentage of genome edited cells by destroying these compatible overhangs, hence mechanically limiting the precise re-joining of designer nuclease-induced DSBs. To this end, we performed in parallel experiments with samples co-transfected with a TREX2 expression construct and the designer nucleases. This however meant that the total amount of designer nuclease constructs has to be reduced to half in order to achieve the recommended 1:1 ratio of designer nuclease : TREX2 and at the same time maintain constant the total amount of plasmid DNA (60μg) delivered.

3.3.5.2.1 Assessment of transfection and targeted disruption efficiencies

The same overall experimental procedure was followed as in the 1st MEL MA821$^{IVSI-110}$ experiment (section 3.3.5.1). Transfection efficiencies (measured by flow cytometry) and targeted disruption of the *HBB$^{IVSI-110}$* locus (determined by T7E1 assay) was assessed 72hr and 5–6 days post-electroporation, respectively. Transfection efficiency was approximately the same for all samples giving approximately 55% GFP$^+$ cells (Figure 61). Targeted disruption was assessed only for the *HBB$^{IVSI-110}$* locus, since the *CCR5* specific TALENs [227] and RGEN used in this experiment are specific for human *CCR5* and not the murine equivalent. In samples without co-transfected TREX2 construct, the T7E1 assay showed targeted disruption of 12.33%, 12.32%

and 4.2% for *HBB*$^{IVSI-110}$ with TAL R1/L1, R1/L2 and RGEN, respectively, whereas in samples including TREX2 disruption dropped approximately by half to 8.37%, 5.9% and 1.36%, accordingly for the TAL R1/L1, R1/L2 and RGEN constructs (Figure 60 & Figure 61).

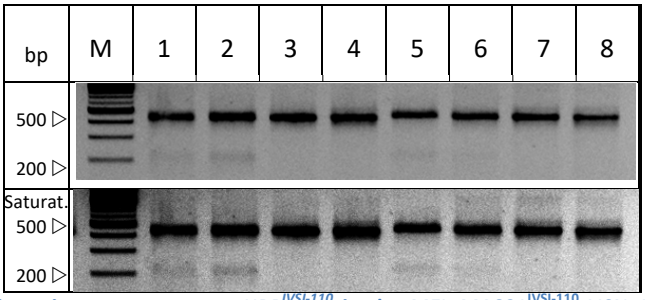


Figure 60: Targeted disruption assessment on *HBB*$^{IVSI-110}$ in the MEL MA821$^{IVSI-110}$ VCN 1.9 cell pool with and without co-transfection with a TREX2 expression construct (2nd Exp).
Following transfection of cells with the designer nucleases either with or without the presence of the TREX2 expression construct, total DNA was subjected to a T7E1 assay and products resolved by agarose gel electrophoresis. Lanes 1, 5: HBB TALEN R1/L1. Lanes 2, 6: HBB TALEN R1/L2. Lanes 3, 7: HBB$^{IVSI-110}$ RGEN. Lanes 4, 5: pUC118 nuclease - free negative (–ve) controls. M: 2log DNA ladder. Uncleaved band 500 bp, cleaved band ~200 bp. Bands were quantified using ImageJ software.

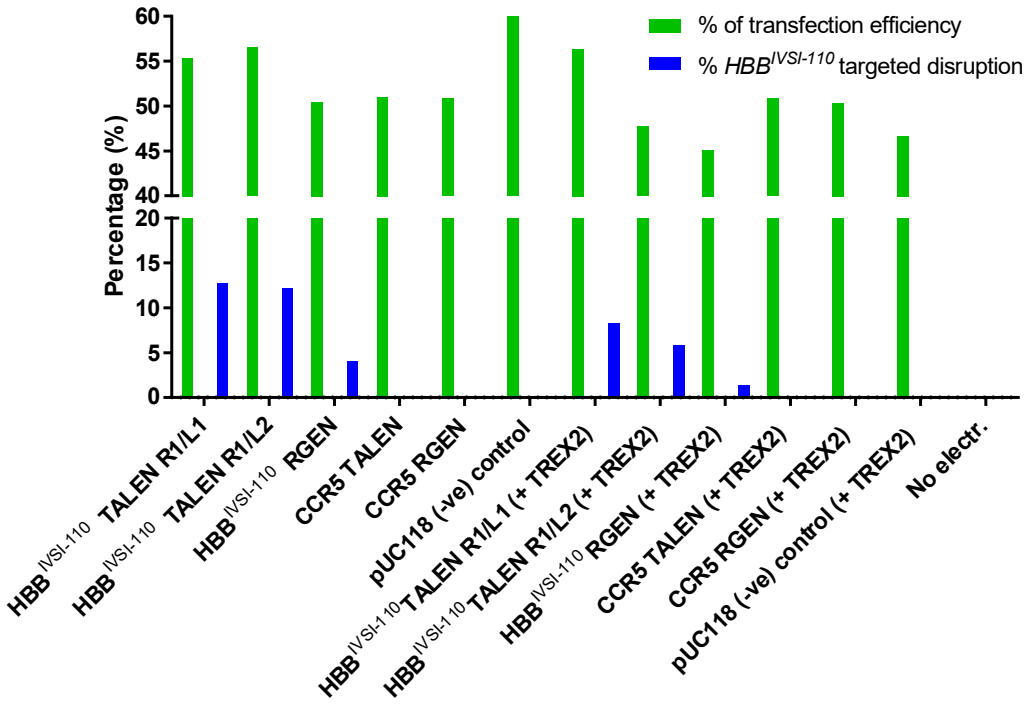


Figure 61: Genome editing of MEL HBB$^{IVSI-110}$ VCN ~1.9 cells with and without co-transfection of the TREX2 expression construct (2nd Exp).
10x10^{6} cells were electroporated with 60 μg of plasmid DNA (50 μg designer nuclease/pUC118 or 25 μg designer nuclease/pUC118 and 25 μg TREX2 and 10 μg CMV GFP reporter). Electroporated cells were culture for 72 hours at hypothermic conditions and then harvested for assessment of transfection efficiency (% of GFP^{+} cells; green bar) via flow cytometry and percentage of genome edited cells in each population measured as targeted disruption of the HBB$^{IVSI-110}$ transgene based on a T7E1 assay (blue bars) is shown. The effect of targeted disruption with *CCR5* specific designer nucleases on the *HBB* locus was not tested. No electroporated cells (No electr.) were used as GFP negative controls for the assessment of transfection efficiency and pUC118 nuclease-free negative control for the assessment of targeted disruption.

3.3.5.2.2 Restoration of correct pre-mRNA splicing

Quantification of the percentage of correctly and aberrantly spliced HBB mRNAs exhibit an increase in the fraction of normal HBB mRNA in samples edited with both TALEN pairs both in the presence or absence of TREX2 both days 3 and 6 of differentiation. In particular, non-edited cells showed ratios of correct/aberrant HBB mRNA of approximately ~0.8 and 0.63 on day 3 and 6, respectively, whereas in the TALEN edited cells with and without TREX2, ratios of c/a increased on average to 2.29, with the maximum restoration detected with the R1/L2 edited samples without TREX2 at 2.66 and 2.79 on days 3 and 6, respectively (Figure 62, top panel). The percentage of *HBB*$^{IVSI\text{-}110}$ RGEN edited samples did not change to a significant degree from the non-edited negative controls, with the exception of the sample co-transfected with TREX2 on day 3 in which the c/a ratio reached 1.2. However, this is most likely in the realms of background noise, since the ratio of the same edited population on day-6 of differentiation returned to level of negative controls. It is worth noting that the samples edited with the *CCR5* specific designer nucleases, are expected to also exhibit similar ratios of correctly/aberrantly spliced mRNAs to that of the non-edited negative controls (Figure 62, top panel). Analysis of the data as percentage change of the two variants compared to the negative (no electroporation) controls, showed more clearly the restoration of correct splicing in cells edited with TALENs. In particular, on day-6 of differentiation of the TALEN R1/L1 edited cells, both with and without the TREX2 expression construct, showed a 95% increase in the correctly spliced HBB mRNA fraction relative to the non-edited control (Figure 62, bottom panel). Thus this 2nd experiment illustrated even more profoundly than the 1st experiment (Figure 58), the higher stability of the normal HBB mRNA compared to the aberrant over time, which is indicated as a progressive increase of the normal mRNA from day-3 to day-6 implying an accumulation of the correct spliced mRNA at higher rates than the aberrant (Figure 62, bottom panel).

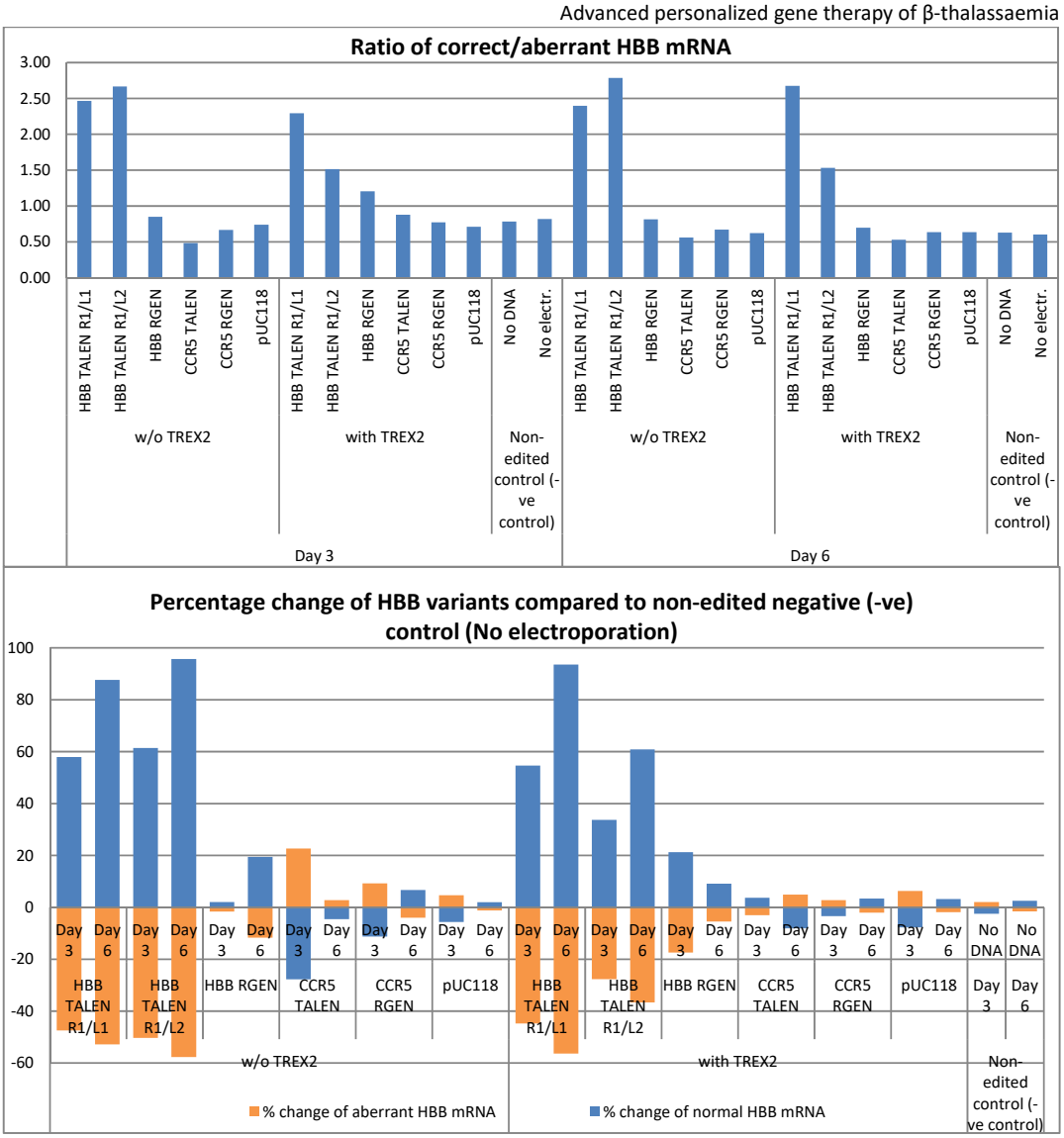


Figure 62: Assessment of the level of restoration of correct splicing of HBB$^{IVSI-110}$ pre-mRNA on day-3 (D 3) and day-6 (D 6) of differentiation of genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cells (2nd Exp).
MEL MA821 HBB$^{IVSI-110}$ cells were genome edited with designer nucleases (HBB TALEN R1/L1, R1/L2, HBB$^{IVSI-110}$ RGEN, CCR5 TALEN and RGEN) with or without TREX2 construct. Genome edited cells were induced differentiated in induction medium (cRPMI suppl. 2% DMSO) and cells were collected for RNA analysis on day 3 and day 6 via multiplex RT-qPCR. Absolute quantities of correct and aberrant HBB mRNA variants were measured using a plasmid-based SC. Top panel: Ratios of normal/aberrant HBB mRNA in the total mRNA population. Bottom panel: Percentage change of correctly (blue bars) and aberrantly (orange bars) spliced HBB mRNAs in genome edited samples relative to non- edited negative (–ve) control levels on day-3 (D 3) and day-6 (D 6) post-erythroid differentiation.

3.3.5.2.3 Restoration of HBB chain production

Besides the extra designer nuclease negative controls and looking at the effect of including the TREX2 expression construct, the differentiation of normal MEL MA821 HBB^{Normal} was also undertaken in order to evaluate the level of correction relative to the “normal" level of correction relative 3.1.2.2, both transgenic MEL cell pools have a similar average VCN of ~1.9. However, cells harbouring *HBB$^{IVSI-110}$* produced only 5.6% of normal HBB chains in *HBB^{Normal}*

containing cells (Figure 26), which as mentioned previously is similar to the expression pattern seen in thalassaemic $HBB^{IVSI-110/IVSI-110}$ patients (~10%) relative to normal individuals [90].

Assessment of correction in edited cells at the level of restoration of HBB chain production as determined by immunoblotting, again showed an obvious increase of the levels HBB chains in samples edited with both TALEN pairs (TALEN R1/L1) co-transfected with TREX2 to 3 times the amount of HBB chains seen in non-edited cells (Figure 63, upper panel). As expected cells treated with *CCR5* - specific designer nucleases, both TALEN and RGEN, gave a similar result to that of the negative control, whereas the MA821 HBB^{Normal} control cells had ~5 times more HBB polypeptide. Re-analysis of the same data expressed as a percentage of the amount of HBB chains in the positive MA821 HBB^{Normal} control, the TALEN pairs without TREX2 had approximately 40% of the normal level of HBB chains, whereas the TALEN R1/L1 with TREX2 reached up to 60% (Figure 64).

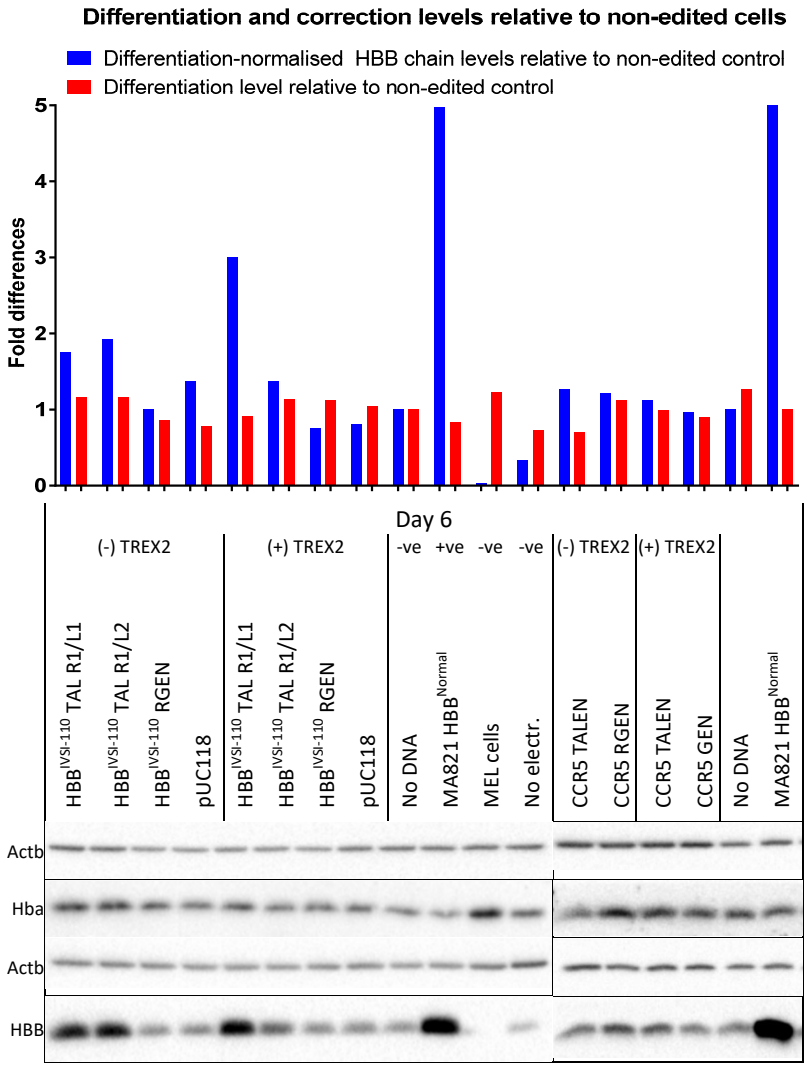


Figure 63: Immunoblot quantification HBB and Hba chains and Actb of protein extracts from day-6 erythroid differentiated, genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cells with and without the TREX2 expression construct (2nd Exp).

Lower panel: MEL MA821 HBB$^{IVSI-110}$ cells were genome edited with designer nucleases (HBB TALEN R1/L1, R1/L2, HBB$^{IVSI-110}$ RGEN, CCR5 TALEN and RGEN) with or without TREX2 construct. Genome edited cells were induced differentiated in induction medium (cRPMI suppl. 2% DMSO) and cells were collected for protein expression analysis on day 6 via western immunoblots analysis. Whole cell extracts from 0.5–1x10^6 cells were probed with antibodies against HBB, Hba and Actb. Actb was used as a protein loading control. Non-edited controls were used as negative control for the assessment of level restoration of splicing of genome edited cells at the protein level. CCR5-edited cells constitute the non-specific designer nuclease negative (–ve) controls. MEL MA821 HBB^{Normal} (VCN 2) positive (+ve) control represents the expected "normal" levels of HBB chains. MEL cells are included as –ve controls for HBB expression. Upper panel: Fold difference of differentiation-normalised HBB chain expression (determined by HBB / Hba chain ratio; blue bars) and differentiation levels (via Hba chain; red bars) compared to non-edited –ve control (no DNA). Band densities were measured using ImageLab.

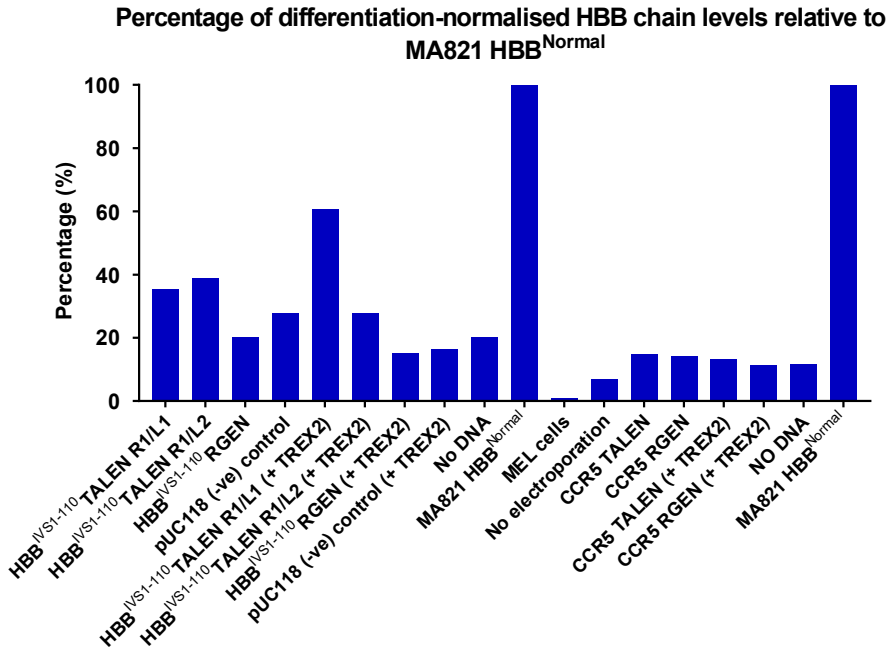


Figure 64: **Percentage of differentiation - normalised HBB chains compared to the "normal" levels in the MEL MA821 HBB^{Normal} VCN 2 positive control cells (2nd Exp)**.
Fold difference of differentiation-normalised HBB chain expression (determined by HBB / Hba chain ratio; blue bars) are illustrated relative to the expression levels of the MEL MA821 HBB^{Normal} positive control. CCR5-edited cells constitute the non-specific designer nuclease negative (–ve) controls. MEL cells are included as –ve controls for HBB expression.

3.3.5.3 *3rd experiment in MEL MA821$^{IVSI-110}$ transgenic cells*

The 3rd experiment using the MEL MA821$^{IVSI-110}$ cell pool platform (average VCN ~1.9) was performed as a further replicate of the above results (sections 3.3.5.1 and 3.3.5.2), and in addition to extend the analysis for characterisation the INDELs, which would have occurred in the edited cell population. Cells were treated with the *HBB$^{IVSI-110}$* specific designer nucleases with and without co-transfection of the TREX2 expression construct. In parallel three nuclease-free negative controls were included as previously described namely, pUC118, no DNA and no electroporation.

3.3.5.3.1 Assessment of transfection and targeted disruption efficiencies

Transfection efficiencies were approximately ~65% for all samples and very high targeted disruption efficiencies were achieved for all samples (Figure 66 & Figure 65). In particular without TREX2, 51.7%, 56.4% and 8.35% targeting efficiencies of *HBB$^{IVSI-110}$* were obtained using TALEN R1/L1, R1/L2 and RGEN, respectively. Inclusion of TREX2 along with these designer nucleases, displayed similar efficiencies of targeted disruption; 49.8 %, 41.5% and

9.5% respectively for TALEN R1/L1, R1/L2 and RGEN (Figure 66 & Figure 65). Thus the inclusion of TREX2 had no enhancing function on targeting disruption of the *HBB*$^{IVSI-110}$ transgene.

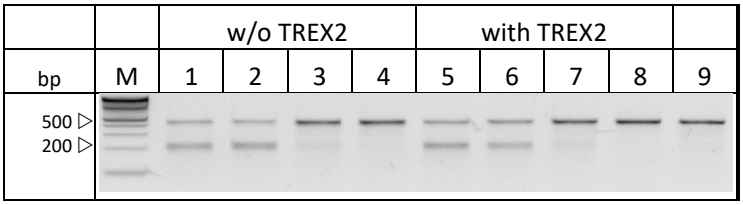


Figure 65: Targeted disruption assessment of the *HBB*$^{IVSI-110}$ transgene on MEL MA821$^{IVSI-110}$ VCN 1.9 cells with and without co-transfection of the TREX2 expression construct (3rd Exp).
Following transfection of cells with the designer nucleases either with or without the presence of the TREX2 expression construct, total DNA was subjected to a T7E1 assay and products resolved by agarose gel electrophoresis. Lanes 1, 5: HBB TALEN R1/L1. Lanes 1, 5: HBB TALEN R1/L1. Lanes 2, 6: HBB TALEN R1/L2. Lanes 3, 7: HBB$^{IVSI-110}$ RGEN. Lanes 4, 8: pUC118 nuclease free negative (–ve) controls. Lane 9: no DNA –ve control. M: 2log DNA ladder. Uncleaved band 500 bp, cleaved band ~ 200 bp. Bands were quantified using ImageJ software.

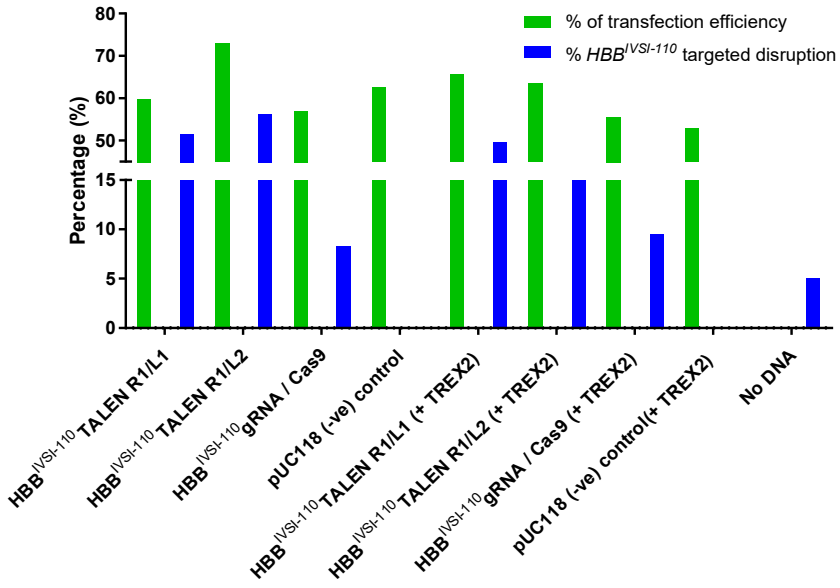


Figure 66: Genomes editing of MEL HBB$^{IVSI-110}$ VCN1.9 cells with and without co-transfection of the TREX2 expression construct (3rd Exp).
10x10^{6} cells were electroporated with 60 µg of plasmid DNA (50 µg designer nuclease/pUC118 or 25 µg designer nuclease/pUC118 and 25 µg TREX2 and 10 µg CMV GFP reporter). Electroporated cells were culture for 72 hours at hypothermic conditions and then harvested for assessment of transfection efficiency (% of GFP^{+} cells; green bar) via flow cytometry and percentage of genome edited cells in each population measured as targeted disruption of the *HBB*$^{IVSI-110}$ transgene based on a T7E1 assay (blue bars) is shown. Electroporated cells without DNA (No DNA) were used as GFP negative controls for the assessment of transfection efficiency and pUC118 nuclease-free negative control for the assessment of targeted disruption

3.3.5.3.2 INDEL characterisation

Further characterisation of INDELs at the site of DNA DSB induction by the designer nucleases rather than just an estimation of the percentage of genome edited cell population by T7E1 assay is required for the better understanding of how exactly the produced INDELs restored correct splicing.

3.3.5.3.2.1 Characterisation of targeted disruptions and frequencies of INDEL size analysis

Targeted disruption efficiencies and frequencies of the different types of INDELs, in terms of size, insertion and deletion were assessed via TIDE analysis [236]. This was achieved by sequencing of PCR products from the genome edited cell population. Due to the presence of different INDELs at the site of the DSB, the sequencing reaction leads to the production of overlapping traces (Figure 67), which can be further analysed via the TIDE web tool (https://tide.nki.nl/). This tool interprets the data from sequencing, as the percentage of targeted disruption and relates this to frequencies of the prominent different size deletions or insertions. As in the T7E1 assay, a non-edited sample is required to be analysed in parallel to be set as a negative control.

We analysed samples treated with the *HBB*$^{IVSI\text{-}110}$ TALENs and RGEN with and without co-transfection of the TREX2 construct. The presence of INDELs in RGEN treated samples were below the level of detection (Figure 68), even though the T7E1 assay showed ~ 10% targeted disruption (Figure 65). The TALEN edited samples showed a relatively high targeted disruption of 21% for both R1/L1, with and without the TREX2, and for R1/L2 pair, 25% and 15% in absence and presence of TREX2, respectively (Figure 68). TIDE analysis gives lower levels (~50%) of targeted disruption compared to the T7E1 assay. As previously reported the TALENs mainly produce deletion rather than insertions following DSB induction [251]. In addition, the presence of the TREX2 does not seem to affect the level targeted disruption or the size of deletions and insertions. Similar frequencies are observed with and without TREX2 in samples edited with same TALEN pair. However, it is noticeable that INDELs in cells edited with the TALEN R1/L1 pair with the longest spacer, 13 bp, tend to produce larger deletions starting from 4 bp up to 8 bp in length, whereas those edited with TALEN R1/L2, spacer 10 bp, the most frequent deletions start at 1 bp and increase up to 9 bp.

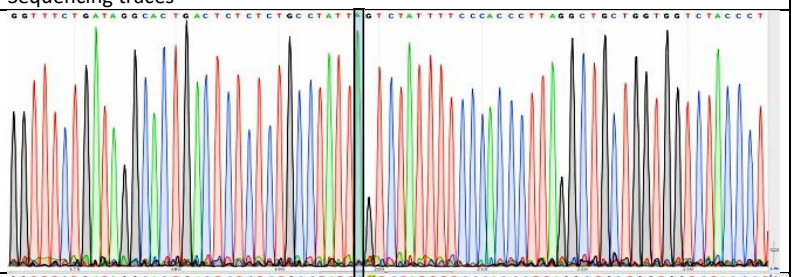
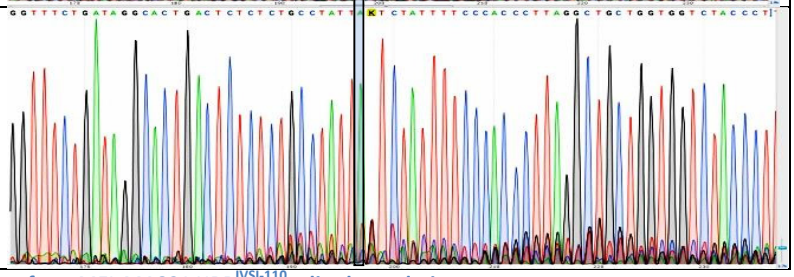
	Sequencing traces
Non-edited control (pUC118)	
HBB$^{IVSI-110}$ TALEN R1/L1	

Figure 67: DNA sequencing traces from MEL MA821 HBB$^{IVSI-110}$ edited population.
DNA sequencing traces of the *HBB$^{IVSI-110}$* transgene region (IVSI-110 (G>A) mutation designated with a blue rectangle) from non-edited MEL cells, pUC118, (Top panel) and cells genome edited with HBB$^{IVSI-110}$ TALEN R1/L1 pair (Bottom panel). The presence of different sizes of INDELs in the genome edited cell population leads to the production of overlapping traces in the sequence read starting to rise a few bases upstream of the HBB$^{IVSI-110(G>A)}$ mutation. In non-edited samples the background levels remain constant through the sequence read. Overlapping DNA sequencing traces from genome edited population can be analysed by Tracking of INDELs by decomposition **(**TIDE) software for the characterisation of targeted disruptions and frequencies of INDEL type and sizes in the population.

INDEL Spectrum

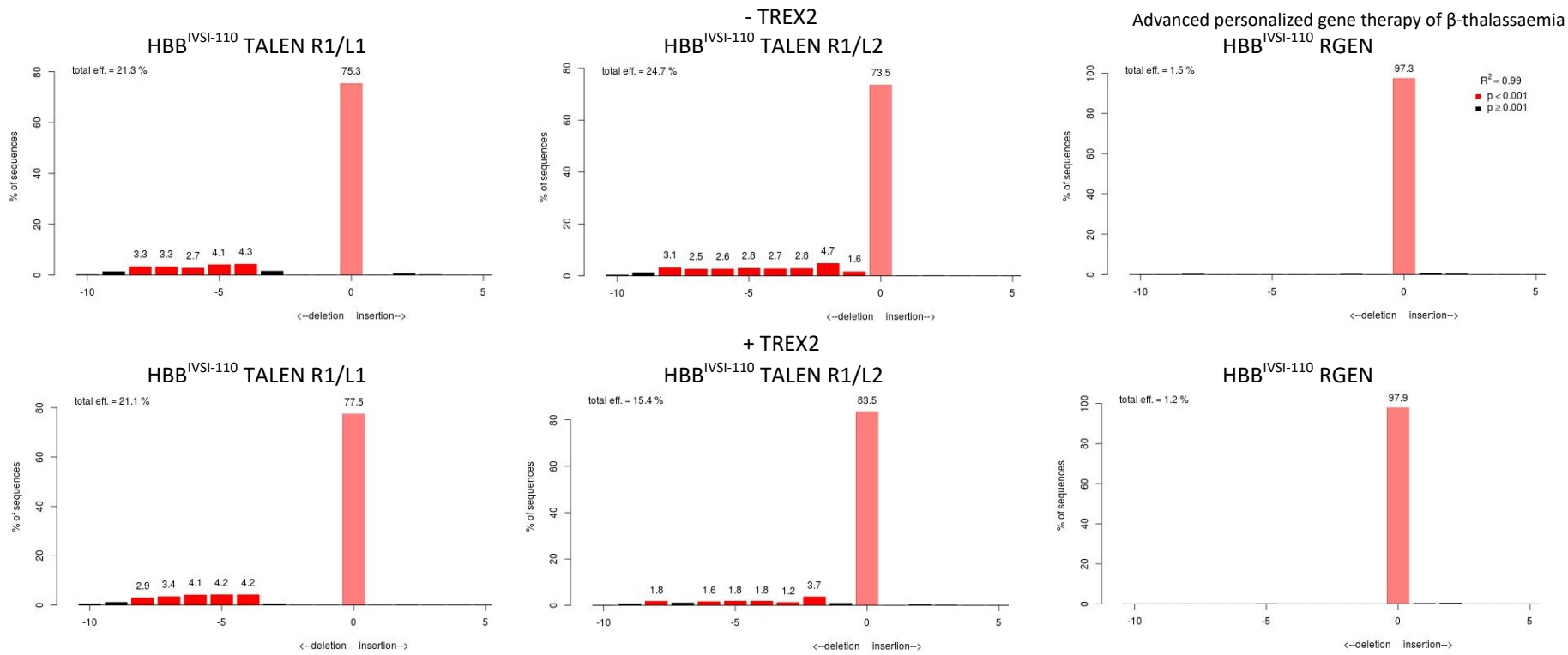


Figure 68. Analysis of INDEL spectrum and frequencies of different types of INDELs in the genome of edited MEL MA821 HBB$^{IVSI-110}$ cell populations by the Tracking of INDELs by decomposition (TIDE) web tool.

The X axis indicates the number of nucleotides deleted (negative axis) or inserted (positive axis), whereas 0 indicates unchanged sequences. The Y axis shows the frequencies of those events. Percentage of genome edited cell population is shown as the total efficiency value. The graph illustrates the TIDE analysis of the MEL MA821 HBB$^{IVSI-110}$ cell population edited by HBB TALEN R1/L1, R1/L2 pairs and RGEN, without -TREX2 and with + TREX2. The main the type of INDELs observed are deletions with no obvious difference observed in the population edited with the same pair of TALENs with and without the TREX2. In addition, comparison of cells edited with the two TALENs indicates that the L1/R1 pair tends to create a deletion starting from 4 bp, whereas with the R1/L1 pair deletions start from 1–2 bp. The RGEN edited population did not show the presence of editing events.

3.3.5.3.2.2 Sanger Sequencing of TOPO clones

Besides the characterisation of INDELs in terms of frequencies and sizes, we analysed INDELs at the level of the DNA nucleotide sequence by Sanger sequencing of TOPO clones harbouring PCR amplicons of the edited *HBB*$^{IVSI-110}$ region. Briefly, PCR products of the *HBB*$^{IVSI-110}$ target locus from edited MEL cell populations were cloned into blunt-end TOPO vectors. A total of 100 single TOPO colonies were picked and sequenced for each editing event induced by TALEN pairs R1/L1 and R1/L2 and RGEN. In addition, 30 clones from the pUC118 nuclease-free negative control were also sequenced as quality controls of the method and as expected no INDELs were detected (data not shown). Clones from edited samples with TREX2 were also picked, however only ~30 clones were sequenced as no obvious differences were detected relative to samples without the co-transfected TREX2 construct (data not shown). Assistance in the inoculation of bacterial cultures and the isolation of DNA for sequencing analysis was gratefully received from Argyro Floga and Petros Ladas. Due to cost and time limitations we prioritised the sequencing of TOPO clones only from edited populations without TREX2. Data from Sanger sequencing showed 41%, 39% and 5% of the clones represented editing events for the R1/L1 and R1/L2 TALEN pairs and RGEN, respectively (Figure 69 & Figure 71). These frequencies correlate well with those observed in the T7E1 assay (51%, 56.4% and 8.39% respectively; (Figure 66). Furthermore, the size of deletions in samples treated with the R1/L1 pair start from 2–3 bp, whereas deletions arising from the R1/L2 pair start as single base disruptions. Graphical representation of the sequencing results as frequencies of different sizes of insertion or deletion display a similar pattern with data obtained from TIDE analysis (Figure 72).

Sequencing analysis showed that not all the INDELs are able to disrupt the aberrant splice site produced by the *HBB*$^{IVSI-110\ (G>A)}$ mutation. In the case of the HBB TALEN R1/L1 edited population, only 14 out of the 40 edited clones directly disrupted the mutation, of which two of them reached close to the normal SA site at the end on intron-1 with one actually destroying this site. The majority of the INDELs occurred upstream of the aberrant splice site (~4 bp away) (Figure 69). Similarly, TOPO clones from the HBB TALEN R1/L2 edited population, only 12 out of 40 showed a disruption of the aberrant splice site, of which one extended into exon II. In addition, the rest of the INDELs tend to be closer to the mutation than the INDELs produced by R1/L1 (Figure 70). Finally, in the case of RGEN edited population, only 5 clones out 100 revealed an editing event and 4 of them disrupted the mutation. This was expected, since the RGEN was designed to create the DSB directly at the site of the mutation, whereas with the TALEN pairs the DSB is expected to occur upstream from the mutation (Figure 47). In general all methods of characterisation of INDELs, in terms of percentage of genome edited

cells and frequencies of the different type and sizes of deletions or insertions, gave similar results with the exception of the targeted disruption efficiency measured by TIDE, which was approximately half of what was observed with the T7E1 assay and TOPO sequencing (Table 11).

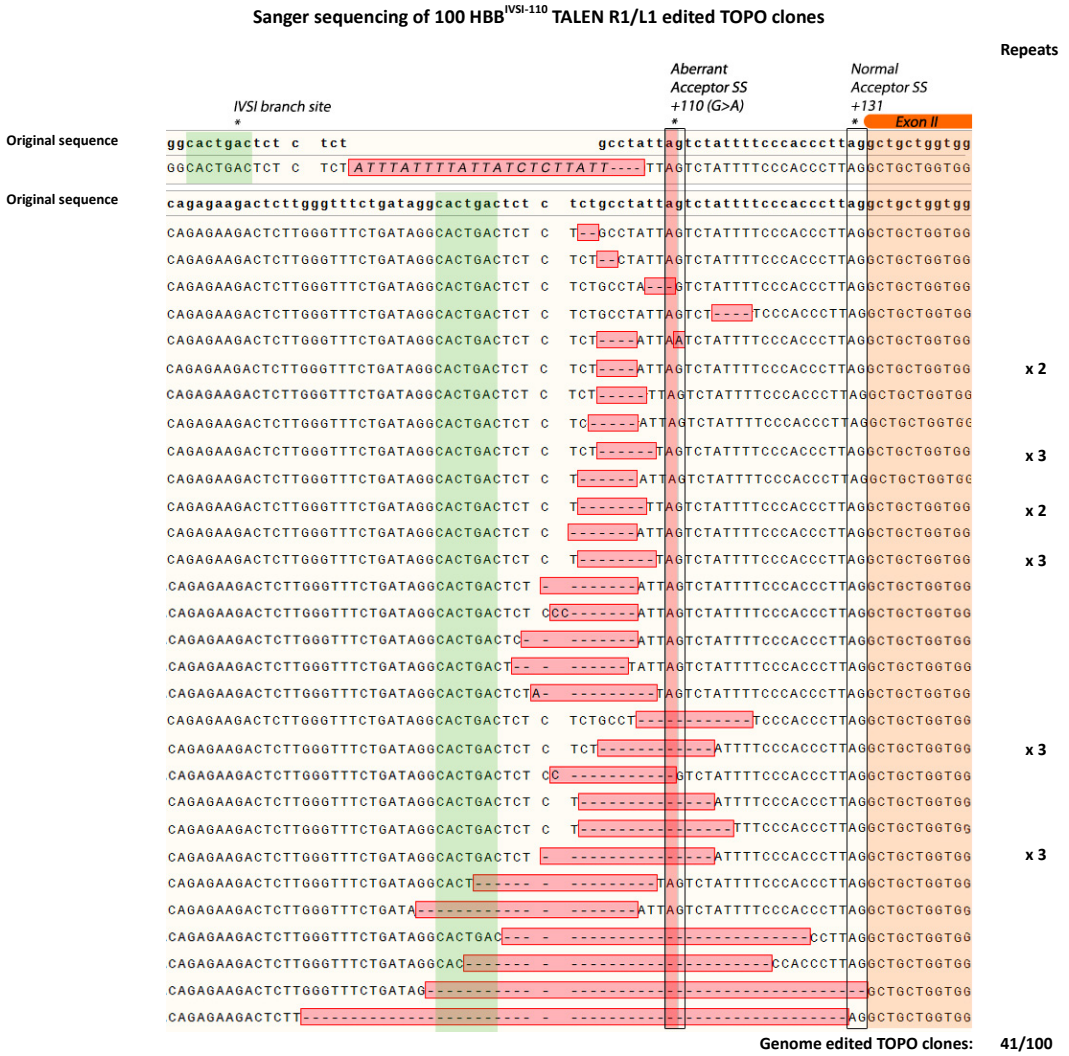


Figure 69: Sanger sequencing of TOPO plasmid vector clones produced from PCR amplification of gDNA from MEL MA821 HBB$^{IVSI-110}$ cells edited by HBB TALEN R1/L1.

The PCR products from the amplification of the HBB$^{IVSI-110}$ region of HBB TALEN R1/L1 genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 population were cloned in TOPO and used for the transformation of TOP10 competent bacteria. 100 colonies were picked and sequenced by Sanger sequencing. Of the 100 TOPO clones sequenced, 41 showed a genome editing event. Clones are aligned primarily based on the size of the INDEL and secondarily based on their proximity to the *HBB$^{IVSI-110\ (G>A)}$* mutation (red vertical box). Orange boxes designate Exon II, green boxes the IVSI branch site and rectangles the abnormal and normal splice acceptor sites (SS). Alignment was performed using SnapGene software.

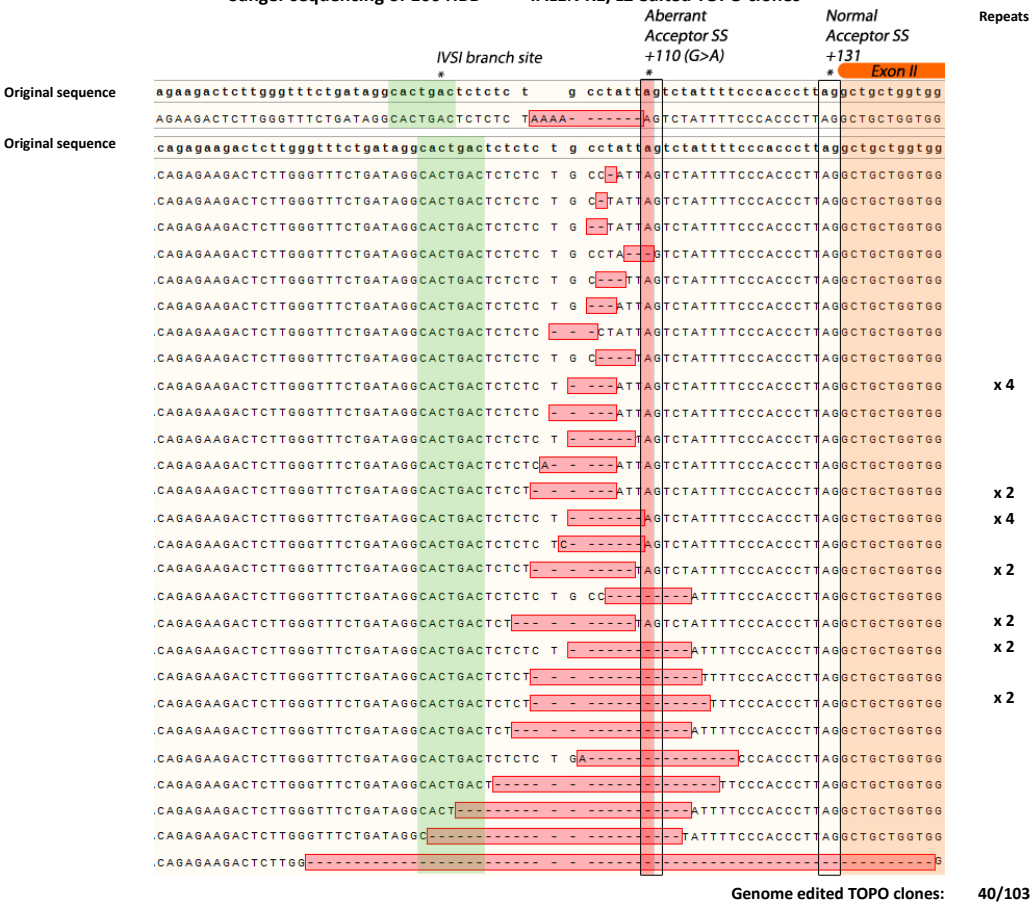


Figure 70: Sanger sequencing of TOPO clones produced by PCR amplification of gDNA from MEL MA821 HBB$^{IVSI-110}$ cells edited with HBB TALEN R1/L2.

The PCR products from the amplification of the HBB$^{IVSI-110}$ region of HBB TALEN R1/L2 genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 population were cloned in TOPO vectors and used for the transformation TOP10 competent bacteria. 103 colonies were picked and sequenced by Sanger sequencing. A total of 103 TOPO clones were sequenced of which 40 showed an editing event had taken place. Clones are aligned primarily based on the size of the INDEL and secondarily based on their proximity to the *HBB$^{IVSI-110\ (G>A)}$* mutation (red vertical box). Orange boxes designate Exon II, green boxes the IVSI branch site and rectangles the abnormal and normal splice acceptor sites (SS). Alignment was performed using SnapGene software.

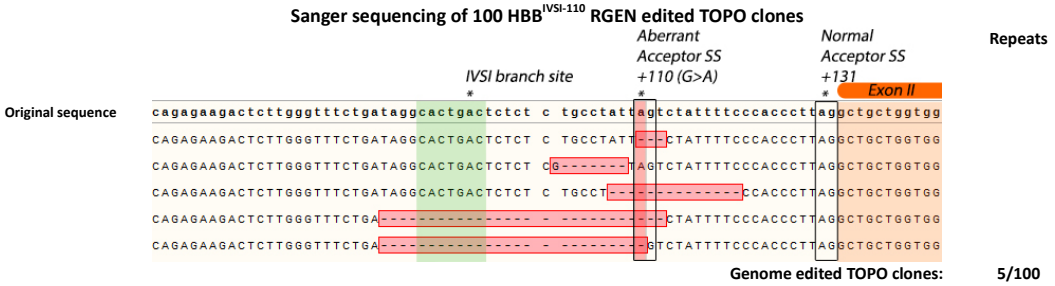


Figure 71: Sanger sequencing of TOPO clones produced by PCR amplification of gDNA from MEL MA821 HBB$^{IVSI-110}$ cells edited with the HBB$^{IVSI-110}$ RGEN.

The PCR products from the amplification of the HBB$^{IVSI-110}$ region of HBB$^{IVSI-110}$ RGEN genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 population were cloned in TOPO vectors and used for the transformation TOP10 competent bacteria. 100 colonies were picked and sequenced by Sanger sequencing. A total of 100 TOPO clones were sequenced revealing only 5 with editing events. Clones are aligned primarily based on the size of the INDEL and secondarily based on their proximity to the *HBB$^{IVSI-110 (G>A)}$* mutation (red vertical box). Orange boxes designate Exon II, green boxes the IVSI branch site and rectangles the abnormal and normal splice acceptor sites (SS). Alignment was performed using SnapGene software.

Table 11.Comparison of the percentage of genome edited cells assessed by different methods.

Quantification of the percentage of genome edited cells from the same populations by different methods namely T7E1 assay, TIDE and Sanger sequencing of individual TOPO clones of PCR products of the targeted *HBB$^{IVSI-110 (G>A)}$* region.

Percentage of genome edited cell population	T7E1 Assay (%)	TIDE (%)	Sanger Sequencing (%)
HBB$^{IVSI-110}$ TALEN R1/L1	51	21.3	41
HBB$^{IVSI-110}$ TALEN R1/L2	56.4	24.7	39
HBB$^{IVSI-110}$ RGEN	8.35	1.5	5

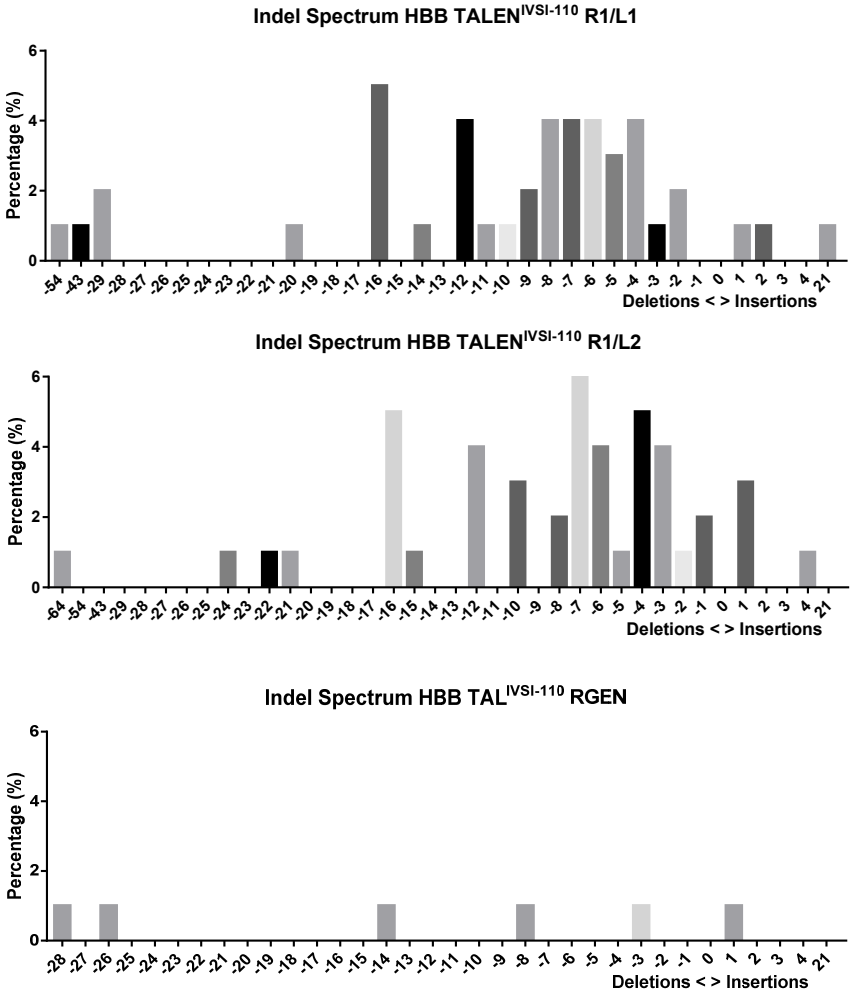


Figure 72: INDELs frequencies detected by Sanger sequencing in TOPO clones.
TIDE-like graphical illustration of the frequencies of different sized INDELs detected by Sanger sequencing of TOPO clones harbouring PCR amplified products of the *HBB*$^{IVSI-110\ (G>A)}$ region derived from MEL MA821 HBB$^{IVSI-110}$ cells edited with the HBB TALEN R1/L1, R1/L2 and RGEN designer nucleases. Graphs of results from TALEN edited samples exhibit similar patterns to those obtained with TIDE analysis. Whereas the TALEN R1/L1 pair tends to produce deletions starting from 4 bp, deletions induced by editing with the R1/L2 TALEN start from 3 bp. However, these data are not conclusive, since the number of edited clones used is relatively low.

3.3.5.3.3 Restoration of correct pre-mRNA splicing

Analysis of the samples at the level of mRNA, showed a major increase of the restoration of correct splicing in samples edited mainly with the two TALEN pairs. In particular, MEL cell non-edited pUC118 samples following day-3 of erythroid differentiation with and without the TREX2 expression plasmid, gave ratios of correct : aberrant mRNA of 0.36 and 0.5 respectively, whereas the corresponding TALEN R1/L1 and R1/L2 edited samples reached a ratio of 2.6 and 2.35 without TREX2 and 3.25 and 2.08 with TREX2, respectively. Accordingly, the same pattern of correction was observed at day-6 of differentiation in the TALEN edited cells (Figure 73, top panel). In addition to the large restoration of correct splicing in cells edited by the TALENs, a partial functional correction of the mutation in samples edited with the RGEN, regardless of co-transfection with TREX2 was also detected for the first time. Specifically, with and without TREX2 at day-3 of differentiation the correct : aberrant splicing ratio increased to 0.78 and 0.7 and on day-6 to 0.57 and 0.51, respectively (Figure 73, top panel). In general, TALEN-edited samples had a greater than 90% change of the fraction of correctly spliced HBB mRNA compared to the non-edited negative control (no electroporation) cell population, whereas the increase in the RGEN treated cells was up to 34% (Figure 73, bottom panel). As shown in previous analyses, the increase in the fraction of correctly spliced HBB mRNA from day-3 to day-6 post-erythroid differentiation relative to non-edited control samples reflects the higher stability of normal HBB mRNA relative to the IVSI-110 aberrant form. This trend was obvious in all the edited samples regardless of the designer nucleases or the addition of TREX2 (Figure 73, bottom panel).

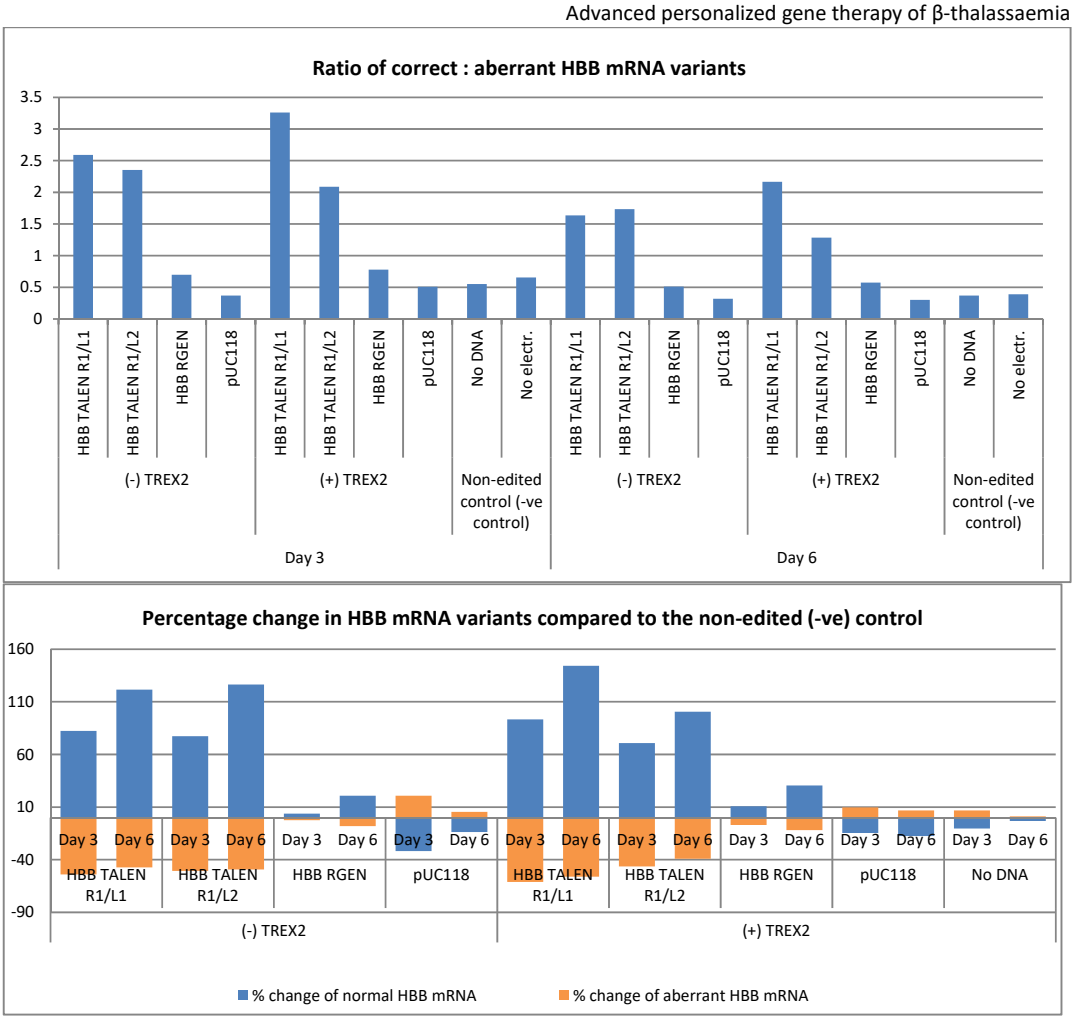


Figure 73: Assessment of level of restoration of correct pre-mRNA splicing at day-3 (D 3) and day-6 (D 6) of erythroid differentiation of genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cells (3rd Exp).
MEL MA821 HBB$^{IVSI-110}$ cells were genome edited with designer nucleases (HBB TALEN R1/L1, R1/L2 and HBB$^{IVSI-110}$ RGEN) with or without TREX2 construct. Genome edited cells were induced differentiated in induction medium (cRPMI suppl. 2% DMSO) and cells were collected for RNA analysis on day 3 and day 6 via multiplex RT-qPCR. Absolute quantities of correct and aberrant HBB mRNA variants were measured using a plasmid-based SC. Top panel: ratios of the correct : aberrant HBB mRNA in the total HBB mRNA population. Bottom panel: percentage change in correctly (blue bars) and aberrantly (orange bars) spliced HBB mRNA in genome edited samples relative to non-edited negative (–ve) control levels (no electroporation) on day-3 (D 3) and day-6 (D 6) of differentiation.

3.3.5.3.4 Restoration of HBB chain production

Assessment of restoration of HBB chain synthesis following genome editing was by immunoblotting. The results show that on day-6 following erythroid differentiation, samples from MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cells edited with TALEN pairs had increased the synthesis of functional HBB chains, reaching up to 8-fold relative to the non-edited (no electroporation) control in the case of the TALEN R1/L1 plus TREX2 edited population (Figure 74). Moreover, the levels of HBB chains reached up to 30% of the amount seen in the normal MA821 HBB^{Normal} control (Figure 75). Remarkably, TALEN-edited samples co-transfected with the TREX2 expression plasmid, gave even higher levels of HBB chain production relative to those without TREX2, although the percentage of genome edited cells was similar at ~45-55% (Figure 66).

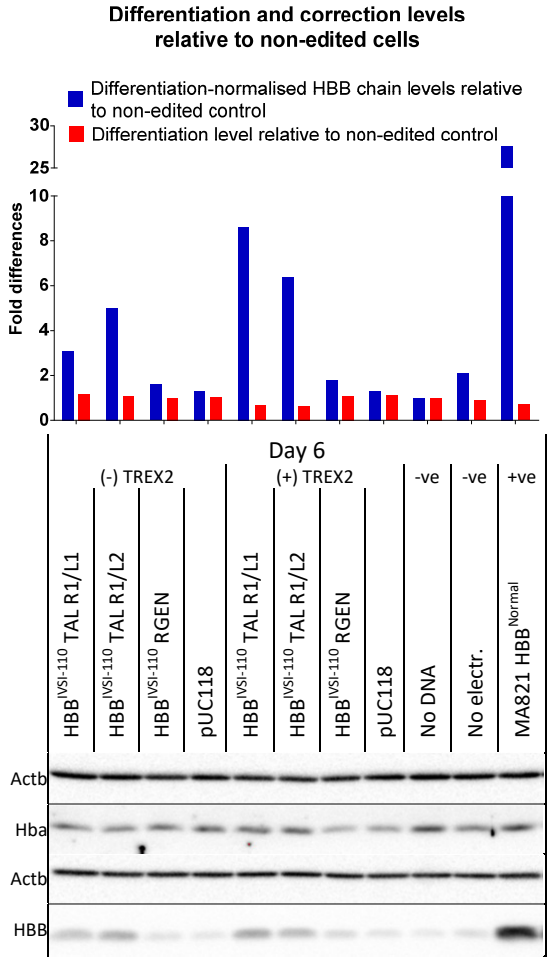


Figure 74: Immunoblot analysis for quantification HBB and Hba chains and Actb in protein extracts at day6 following erythroid differentiation of genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cells with and without the TREX2 construct (3rd Exp).

Lower panel: MEL MA821 HBB$^{IVSI-110}$ cells were genome edited with designer nucleases (HBB TALEN R1/L1, R1/L2 and HBB$^{IVSI-110}$ RGEN) with or without TREX2 construct. Genome edited cells were induced differentiated in induction medium (cRPMI suppl. 2% DMSO) and cells were collected for protein expression analysis on day 6 via western immunoblots analysis. Whole cell extracts from 0.5–1x10^{6} cells were probed with antibodies against HBB, Hba and Actb. Actb was used as a protein loading control. Non-edited controls were used as negative control for the assessment of level restoration of splicing of genome edited cells at the protein level. MEL MA821 HBB^{Normal} (VCN 2) positive (+ve) control represents the expected "normal" levels of HBB chains. MEL cells are included as –ve controls for HBB expression. Upper graph: fold difference of differentiation-normalised HBB chain expression (via HBB / Hba chain ratio; blue bars) and differentiation levels (via Hba chain; red bars) compared to non-edited negative (–ve) control (no DNA) sample. Actb was used as a protein loading normalisation control. Band densities were measured using ImageLab.

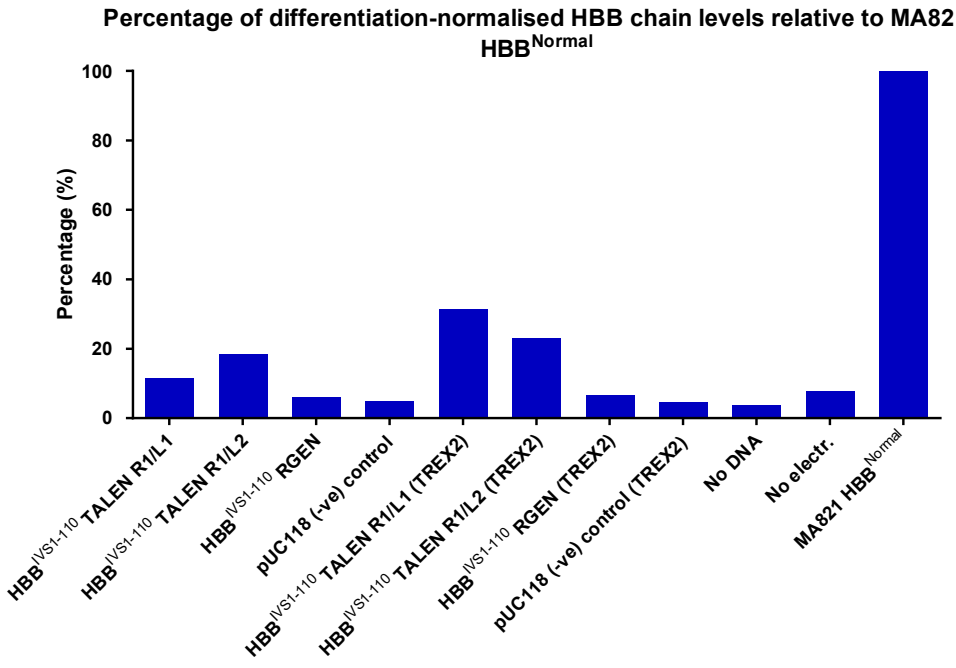


Figure 75: Percentage of MEL cell differentiation normalised HBB chains compared to the "normal" levels in MEL MA821 HBB^{Normal} VCN 2 positive (+ve) control cells (3rd Exp).
Fold difference of differentiation-normalised HBB chain expression (determined by HBB / Hba chain ratio; blue bars) are illustrated relative to the expression levels of the MEL MA821 HBB^{Normal} positive control. MEL cells are included as –ve controls for HBB expression.

3.3.5.4 Summarizing restoration of splicing in MEL cells

The aim of this phase of the project was to characterise the type of INDELs produced after genome editing of MEL MA821 HBB$^{IVSI-110}$ VCN ~1.9 cells with the two HBB TALEN pairs, R1/L1 and R1/L2, and RGEN by three different methods, T7E1 assay, TIDE and Sanger sequencing of TOPO clones. In general, TALEN pairs were more efficient than the RGEN at producing the desired editing event at the MA821 HBB$^{IVSI-110}$ transgene locus with the addition of TREX2 not enhancing the level of disruption to a noticeable degree. Genome editing with the TALEN pairs R1/L1 and R1/L2 mainly produced deletions rather than insertions, in accord with results of previous studies [181]. The R1/L1 TALEN pair induced deletions that tended to start from 3–4 bp, whereas deletions induced by the R1/L2 TALEN started from 1–2 bp. Both TALEN pairs target the same region with the L1 monomer designed to bind 3 bp upstream from L2, while both have the same R1 monomer. One possible explanation of the larger deletion produced by the L1/R1 pair is that the spacer is larger, 13 bp, and therefore the *Fok*I cleavage domains are activated and create nicks which are slightly further than those produced with the R1/L2 pair, 10bp spacer. This probably creates larger overhangs, which are more permissive to the removal of genetic material upon repair by the NHEJ pathway than the slightly shorter overhangs. A similar suggestion has been proposed in an attempt to explain the difference in

mutation patterns produced in samples edited with TALENs and ZFNs, even though both use the same *Fok*I cleavage domain [181]. This previous study showed that TALENs mainly produced deletions (89% deletions versus 1.6% insertions), whereas ZFN-induced insertions and deletions are more evenly distributed. The authors believe that the observed different mutation patterns between TALENs and ZFNs are due to differences in the length of spacers, in particular 12–21 bp and 5–6 bp, respectively. The larger spacer of TALENs possibly gives rise to a more heterogeneous set of overhangs than the defined 4 or 5 bp overhangs produced with ZFNs [181]. Importantly, further to INDEL characterization, the work presented here established proof of principle at a transcriptional and translational level with results clearly showing disruption of the IVSI-110 aberrant SA site and the adjacent region, especially with our most efficient designer nucleases, TALENs R1/L1 and R1/L2. As a result correct pre-mRNA splicing was restored, which was measured as an increase in the percentage of normal HBB mRNA and HBB chain production by ~100% and up to 8 fold, relative to non-edited cells, respectively. Statistical analysis of the restoration of pre-mRNA splicing showed a significant increase of the ratio of correct to aberrant HBB mRNA in samples edited with HBB TALEN R1/L1, R1/L2 to 2 and ~1.5 on day-3 and day-6 of MEL erythroid differentiation, respectively whereas ratios in non-edited control remained below one (Figure 76). Samples co-transfected with the TREX2 expression construct along with the designer genome editing nucleases displayed similar or even higher ratios of correct : aberrant HBB mRNA. However, more experiments need to be performed to test if the difference with and without TREX2 is statistically significant. Estimation of the expression levels of MEL cell differentiation-normalised HBB chains in edited subpopulations from all the experiments, showed an overall increase of the expression in all the edited cells regardless of HBB specific designer nuclease (TALEN or RGEN). Samples edited with HBB TALEN R1/L1 and TREX2 reached on average a 20±6.16-fold (n=2) higher increase in HBB expression than the non-edited negative controls, whereas the R1/L2 had the highest fold increase in edited samples without TREX2, 5.95±3.52 (n=4) (Figure 77). Statistical analysis was performed only in samples repeated four times in which although an increases was observed, the difference was not statistically significant compared to the No DNA negative control (Figure 77).

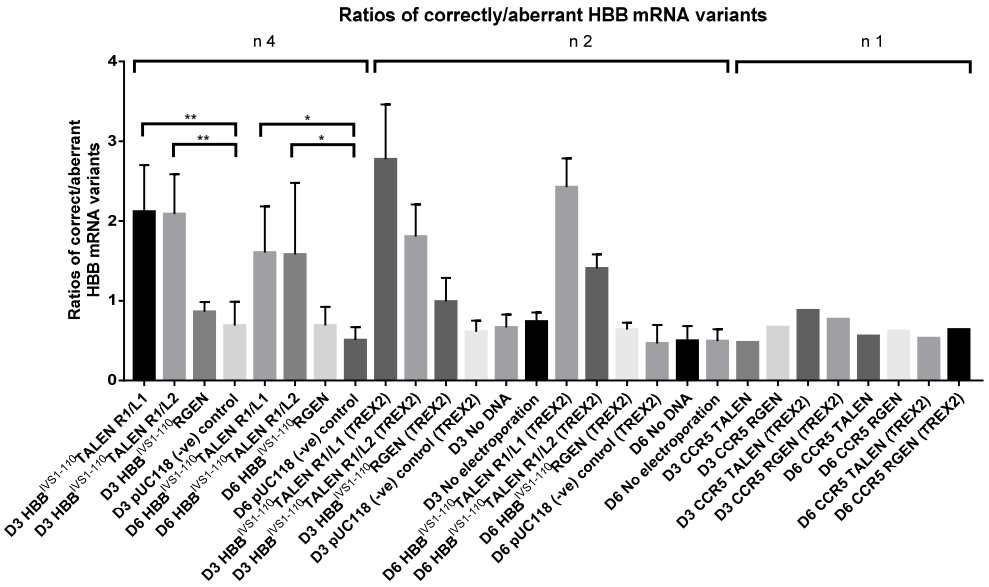


Figure 76: Restoration of correct HBB mRNA splicing in MEL pools on days 3 and 6 of erythroid differentiation across all experiments.

A comparison of the ratios of fractions of correctly spliced : aberrantly spliced HBB mRNA variants was conducted. Statistical analysis was used only in samples repeated four times (n=4) and edited samples were compared at the same day of differentiation as nuclease-free negative (–ve) control (pUC118) samples. Samples repeated twice (n=2) with the HBB specific-designer nucleases and TREX2 and those performed only once (n=1) *CCR5* specific-designer nucleases with and without TREX2 are illustrated without testing for statistical significance differences.

Statistical analysis was performed by one-way ANOVA.

Statistical analysis:

DAY 3 HBB TAL R1/L1	**	– *p* value 0.0011
DAY 3 HBB TAL R1/L2	**	– *p* value 0.0014
DAY 6 HBB TAL R1/L1	*	– *p* value 0.0159
DAY 6 HBB TAL R1/L2	*	– *p* value 0.0193

* $p < 0.05$, ** $p < 0.01$*** $p < 0.001$, **** $p < 0.0001$.

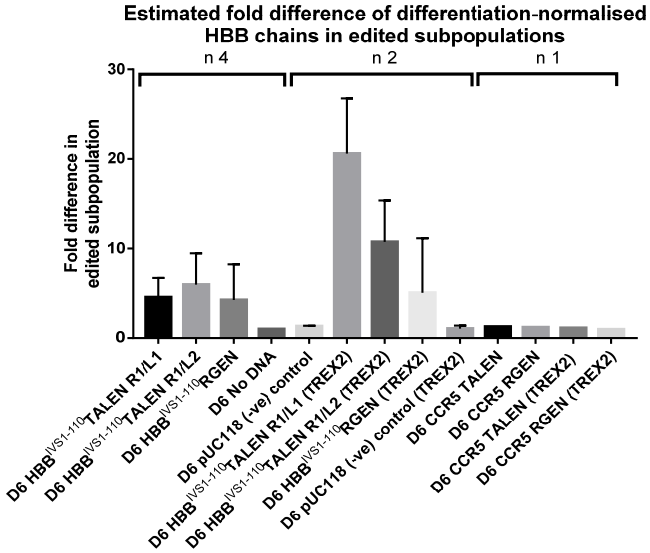


Figure 77: Estimated fold difference of HBB/Hba chain ratio in MEL cells edited with HBB TAL R1/L1, R1/L2, and RGEN.

Transfections of MEL MA821 HBB$^{IVSI-110}$ with designer nucleases were with or without TREX2 and analysed as cell pools at day 6 of induced erythroid differentiation relative to nuclease-free negative (–ve) controls (No DNA). Statistical analysis was performed only in samples repeated four times (n=4), in which there is an increase of up to 6-fold in the R1/L2 samples, but which owing to high intra-group variation does not reach statistical significance compared to the No DNA –ve control. Samples repeated twice (n=2) with *HBB* targeting specific-designer nucleases and TREX2 and those performed only once (n=1) with *CCR5*-targeting specific-designer nucleases with and w/o TREX2 are illustrated w/o testing for statistical significance. Statistical analysis was performed by one-way ANOVA.

3.3.6 *Functional characterization of INDELs in clonal MEL cells*

The aim of this phase of the project was to attempt to correlate specific INDELs with the level of expression at an mRNA and protein level, which will assist in understanding how the splicing is affected and identify specific alternations of the *HBB*$^{IVSI-110}$ transgene that restores splicing. To this end, genome editing of MEL MA821 HBB$^{IVSI-110(G>A)}$ cell clone (VCN 1) was undertaken with HBB$^{IVSI-110}$ TALEN R1/L1 and R1/L2 pairs and RGEN without co-transfection of the TREX2 expression construct. In parallel, nuclease-free negative controls, pUC118 and no DNA were included in the experiment. Transfection efficiencies were relatively low being 13%, 17%, 4.42% and 6.52% for HBB$^{IVSI-110}$ TALEN R1/L1 and R1/L2 pairs, RGEN and pUC118, respectively (Figure 78). Targeted disruption based on a T7E1 assay was 31.5%, 7.7% for HBB$^{IVSI-110}$ TALEN R1/L1 and R1/L2 pairs and 3% for the RGEN (Figure 78 & Figure 79).

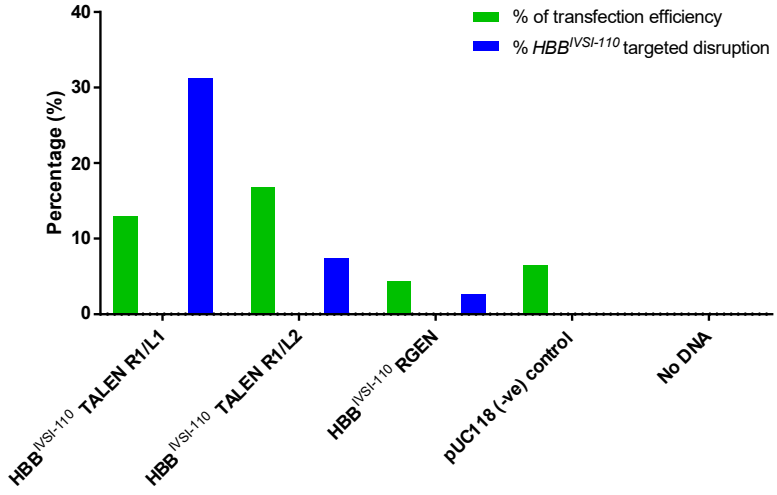


Figure 78: Genome editing of the MEL HBB$^{IVSI-110}$ VCN 1 clonal cell line.
10x10^6 cells were electroporated with 60 µg of plasmid DNA (50 µg designer nuclease/pUC118 and 10 µg CMV GFP reporter). Electroporated cells were cultured for 72 hours at hypothermic conditions and then harvested for assessment of transfection efficiency (% of GFP$^+$ cells; green bar) via flow cytometry and percentage of genome edited cells in each population measured as targeted disruption of the *HBB*$^{IVSI-110}$ transgene based on a T7E1 assay (blue bars) is shown. Electroporated cells without DNA (No DNA) were used as GFP negative controls for the assessment of transfection efficiency and pUC118 nuclease-free negative control for the assessment of targeted disruption

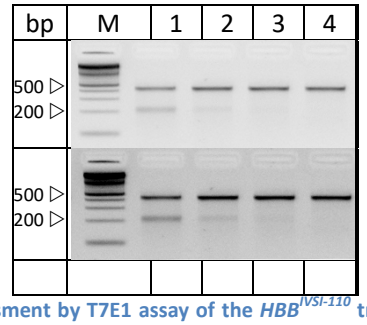


Figure 79: Targeted disruption assessment by T7E1 assay of the *HBB*$^{IVSI-110}$ transgene on MEL MA821$^{IVSI-110}$ VCN 1 clone cell line.
Following transfection of cells with the designer, total DNA was subjected to a T7E1 assay and products resolved by agarose gel electrophoresis. Lane 1: HBB TALEN R1/L1; lane 2: HBB TALEN R1/L2; lane 3: HBB$^{IVSI-110}$ RGEN; lane 4: pUC118 nuclease free –ve controls. M: 2log DNA ladder. Uncleaved band 500 bp, cleaved band 200 bp. Bands were quantified using ImageJ software.

Genome edited populations were seeded in 200–250 wells of 96-well plates at a dilution to give a single cell per well. Since targeted disruption of cells edited with the $HBB^{IVSI-110}$ RGEN was very low (3%), twice as many wells were seeded with this pool of cells than with the rest of the samples (~400–500 wells). Wells with single cell clones were picked, expanded, and gDNA extracted. PCR amplified products from the gDNA corresponding to the IVSI-110 mutant site of the MA821 $HBB^{IVSI-110(G>A)}$ transgene were sequenced to identify the nature of any INDELs. The results of the single clone sequencing showed 19 out of 69 (27.9%) and 20 out of 83 (24.1%) were genome edited with $HBB^{IVSI-110}$ TALEN R1/L1 and R1/L2 pairs respectively. Contrastingly, all the sequenced RGEN treated clones were non-edited, even though approximately twice the number of clones (143) were picked and sequenced compared to the TALEN-edited samples. Clones transfected with pUC118 (24) were also picked and sequenced in parallel with the rest of the samples and as expected none of them showed any genome editing. Until characterisation of genome edited clones by Sanger sequencing, all clones were cryopreserved. Upon characterisation, genome edited clones were thawed, expanded and re-sequenced to confirm their identity. Out from 39 clones, one (A1 #49) failed to expand, two (A1 #7 and A1 #57) gave overlapping sequencing traces indicating that these were not clonal but contained a second non-edited cell population growing in parallel with the edited population, and two (A1 #32 and A2 #65) were not edited, which meant that the wrong clones were recovered, probably due to mislabelling. However, both non-edited clones were expanded and used as extra negative controls (Figure 81). Clones are aligned from top to bottom, based initially on the size of the INDEL (small to large) and subsequently according to their distance from the $HBB^{IVSI-110(G>A)}$ mutation.

3.3.6.1 *Functional correlation of INDELs in genome edited MEL cell clones at a transcriptional and translational level*

All edited and non-edited control MEL MA821 $HBB^{IVSI-110(G>A)}$ cell clones were induced to undergo erythroid differentiation. Cells were harvested for total RNA and protein extraction at days 3 and 6 of differentiation, respectively.

3.3.6.1.1 Restoration of correct pre-mRNA splicing

As described in previous genome editing experiments on transgenic MEL MA821 $HBB^{IVSI-110}$ average VCN ~1.9 cell pool, the percentage of correct and aberrant HBB mRNA variants were quantified using a multiplex RT-qPCR (section 3.1.2.1). In addition, quantification of expression of total HBB variants normalised with respect to the Hba mRNA was compared to non-edited MEL MA821 $HBB^{IVSI-110}$ VCN 1 clone, using the ΔΔCt method. Combining the absolute quantities of the two HBB variants measured by multiplex RT-qPCR and the different expression levels of

the differentiation normalised total HBB mRNA relative to non-edited clones, enabled calculation of the absolute quantities of the two HBB mRNA variants normalised for differentiation in each clone (Figure 81). These data are very informative, especially for clones which contain similar percentages of correct and aberrant HBB mRNAs but with the different expression levels.

For example, A2 #56, A2 #49, A1 #4, A1 #14 and A1 #8 are clones with different modifications on their genome. However, they produced similar percentage of aberrant and correctly spliced HBB mRNA on day-3 of differentiation, (10.39±2.20% aberrant). Even though the fractions of the two HBB variants are similar, the normalised quantities differ; in particular, compared to the clone with the lowest quantities of normal HBB mRNA, A1 #4, the A2 #56 and A1 #14 have slightly higher quantities, 2 and 5 fold higher respectively. In contrast, clones A2 #49 and A1 #8 are 16.4 and 26.18 times higher than A1 #4. In addition, clones with the same deletion, probably derived from the same clone, A1 #22, A1 #31, A2 #53, A2 #3 and A2 #15, have a similar percentage of HBB mRNA variants, (17.23±6.06% aberrant). A comparison of the normalised quantities of the normal HBB mRNAs with the clone with the lowest quantities, A2 #15, a 2.09, 5.12, 5.13 and 3.47 fold-difference for A1 #22, A1 #31, A2 #53 and A2 #3 respectively are observed. These fluctuations in the quantities of correctly spliced HBB mRNA should be taken into account as they can explain differences seen at a translational level. In the case of differences in clones with the same deletion, where an increase of up to 5-fold in normal HBB mRNA was detected, this could be explained as variation in the extent of erythroid differentiation. Normal HBB mRNA is relatively more stable than the aberrantly spliced IVSI-110 variant and accumulates upon differentiation (% change from day 3 to day 6). Therefore the clones that are able to differentiate more effectively in all probability contain more normal HBB mRNA.

Furthermore, different clones which exhibit large differences in the normalised quantities of correct HBB mRNA, even though they have similar percentages of correct and aberrantly spliced HBB mRNAs, could be explained as an alternation of the overall rate of splicing due to the specific INDELs they contain. The importance of combining the absolute quantities of the two variants with the relative quantification of the differentiation-normalised total HBB mRNA variants is also shown by an extreme example in the case of the two clones (A2 #43 and A2 #75) with large deletions in which the normal SA site and part of Exon 2 was disrupted. Both clones show 100% of normal HBB mRNA, as the result of the detection the HBB^{Normal} mRNA curves in the multiplex RT-qPCR, in only one of the triplicate samples at high Ct values of 39.8 and 36.3, respectively, whereas the rest remained undetermined. However, when the two are seen as normalised quantities than just as percentages of correct and aberrant mRNAs is

obviously that the calculated 100% of correctly spliced mRNA is misleading since based on the quantities of the transcripts it is obvious that the expression of both variants is abolished.

In general, all clones in which the *HBB*$^{IVSI\text{-}110\ (G>A)}$ mutation is disrupted, correct splicing is restored, except in the case of clones A2 #43 and A2 #75, which as mentioned above harbour a deleted normal SA site. However, restoration of splicing was also seen with the disruption of the region upstream of the *HBB*$^{IVSI\text{-}110(G>A)}$ mutation without affecting the aberrant SA site. In fact, it was noticed that in general, deletions larger than 5 bp (starting from A2 #63) completely restore pre-mRNA splicing. It is worth noting the unique case of complete restoration of correct splicing in the clone with the smallest INDEL namely A1 #43, which contains a 1 bp (T) deletion. These data suggest that disruption of the region upstream from the SA site is sufficient to restore partially or even completely correct splicing. One explanation, on which our own genome editing hypothesis is also based, is that upstream from each SA site (in this case aberrant SA site) there are regulatory sequence (such as the PYT) which affect the efficient recruitment and assembly of pre-spliceosome at the specific site [77]. Therefore, we believed that the disruption of those elements is able to alter the dynamic of splicing in favour to the non-affected normal SA site.

From the 34 genome edited MEL clones, 22 exhibited complete restoration of correct splicing (0.13±0.32% aberrant) with similar normalised quantities of correct HBB mRNA. In particular, the fold differences of normalised quantities of correct HBB mRNA relative to the clone (A1 #1) with the lowest amount were within the range (1–7.39 fold), which it could be explained due to difference in the degree of differentiation, . Note that the clone which had the highest fold difference of 7.39 is A2 #38, which is a clone with the same editing modification as A2 #61 that exhibits a 3.4 fold difference from A1 #1.

3.3.6.1.2 Restoration of HBB chain production

Protein extracts from all the 6-day differentiated genome edited clones were analysed by immunoblotting for detection and quantification of HBB and Hba chains normalised against Actb as a protein loading control. Non-edited pUC118 clones, Mock A and B, were also included as negative controls and references for the comparison on different immunoblots (Figure 80). As described in previous experiments, band densities were quantified, from which the degree of erythroid differentiation is measured as the amount of Hba chains (Figure 80, top panel) and restoration of differentiation-normalised HBB chain production (HBB / Hba chains) both relative to non-edited controls Mock B (Figure 80, bottom panel). As expected clones showed different patterns of HBB chains production, whereas the level of differentiation in the majority of the clones was similar with a 0.6–1.4 fold difference.

However, extreme values were also detected, such as with clones A1 #22, A1 #14 (0.1 fold) and A3 #3 (2.1 fold). Extreme values of differentiation can lead to over- or under- estimation of restoration splicing after normalising for differentiation.

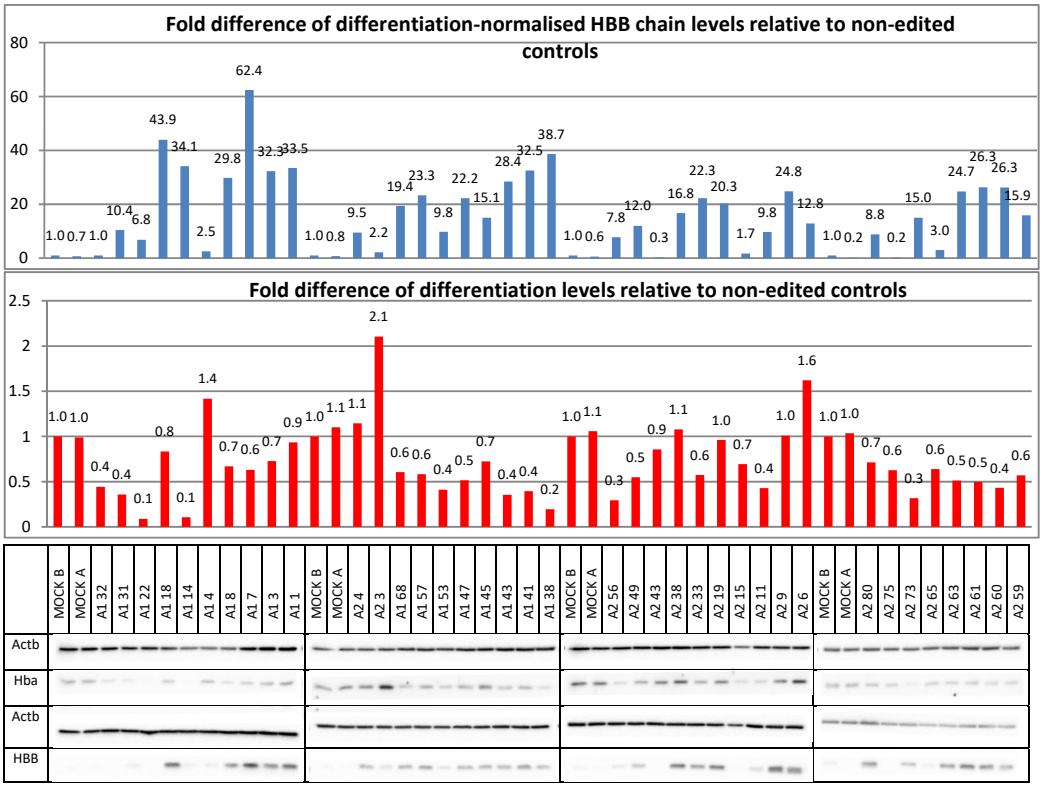


Figure 80: Assessment of HBB expression and differentiation in genome edited MEL MA821 HBB$^{IVSI-110}$ clones relative to non-edited.

Confirmed MEL MA821 HBB$^{IVSI-110}$ genome edited clones by Sanger sequencing, were induced differentiated in induction medium (cRPMI suppl. 2% DMSO) and cells were collected for expression analysis at the RNA and protein level on day-3 and day-6, respectively. Protein expression analysis was performed via western immunoblots analysis. Whole cell extracts from 0.5–1x10^6 cells were probed with antibodies against HBB, Hba and Actb. Actb was used as a protein loading control. Clones treated with HBB TALEN pair R1/L1 or R1/L2 are referred to as A1 or A2 # clones. Quantification of protein bands for HBB and Hba chains normalised for protein loading by reference to Actb from protein extracts of MEL MA821 HBB$^{IVSI-110}$ edited clones on day-6 of erythroid differentiation. MOCK A and B are non-edited clones treated with pUC118 and are used as negative (–ve) controls and as a reference for indirect comparison of clones on different immunoblots for the assessment of level restoration of splicing of genome edited clones at the protein level. A1 #32, A1 #65 are non-edited clones treated with the corresponding TALEN pairs. Red bars indicate the level of differentiation (via Hba chains) and restoration of HBB chain production as blue bars. The fold difference of differentiation normalised HBB chains (HBB / Hba chains) both relative to non-edited control Mock B is shown. Band densities were measured using ImageLab.

Correlation of the protein data from day-6 erythroid differentiated edited MEL MA821 HBB$^{IVSI-110}$ VCN 1 cell clones with specific INDELs, revealed that most of them are in agreement with the data from the correlation of the same INDELs with the HBB mRNA pattern at day-3 of differentiation (Figure 81). Initially, two of the clones (A1 #22, A2 #3) with the same editing had very different degrees (0.09 and 2.1 fold respectively) of differentiation, relative to non-edited clone, whereas the remaining clones (A1 #31, A2 #53, A2 #15) were similar at 0.36, 0.41 and 0.69 respectively. Differentiation-normalised HBB chain levels in clones A1 #31 and A2 #53 were similar at 10.4 and 9.76, whereas for the more differentiated A2 #15 clone it was 1.74

times that of the non-edited control. On average this specific deletion in *HBB* Intron 1 led to a 7.31±3.9 fold increase in HBB chain production compared to non-edited samples, even though approximately 16.8 % HBB mRNA is still aberrant (Figure 81) and that the average normalised quantities of normal HBB mRNA is only ~4% compared to the clones with the full restoration (99–100%). In addition, the identical A2 #38 and A2 #61 clones which exhibited 99–100 % restoration of splicing, showed a 16.8 and 26.3 fold increase of HBB/Hba levels and 1.08 and 0.5 fold difference of differentiation relative to the non-edited control. As expected clones with large deletions (A2 #43 and A2 #75) extending into *HBB* Exon II, did not show production of HBB chains (0.3 and 0.2 fold respectively), even though they were well differentiated (0.86 and 0.63 fold respectively). In general, as was also seen at the mRNA level, increase levels of differentiation normalised HBB chains was observed in clones starting from a 5 bp deletion, with the exception of the 1 bp deletion in clone A1 #43, where we detected up to a 28-fold increase in HBB chains relative to non-edited controls.

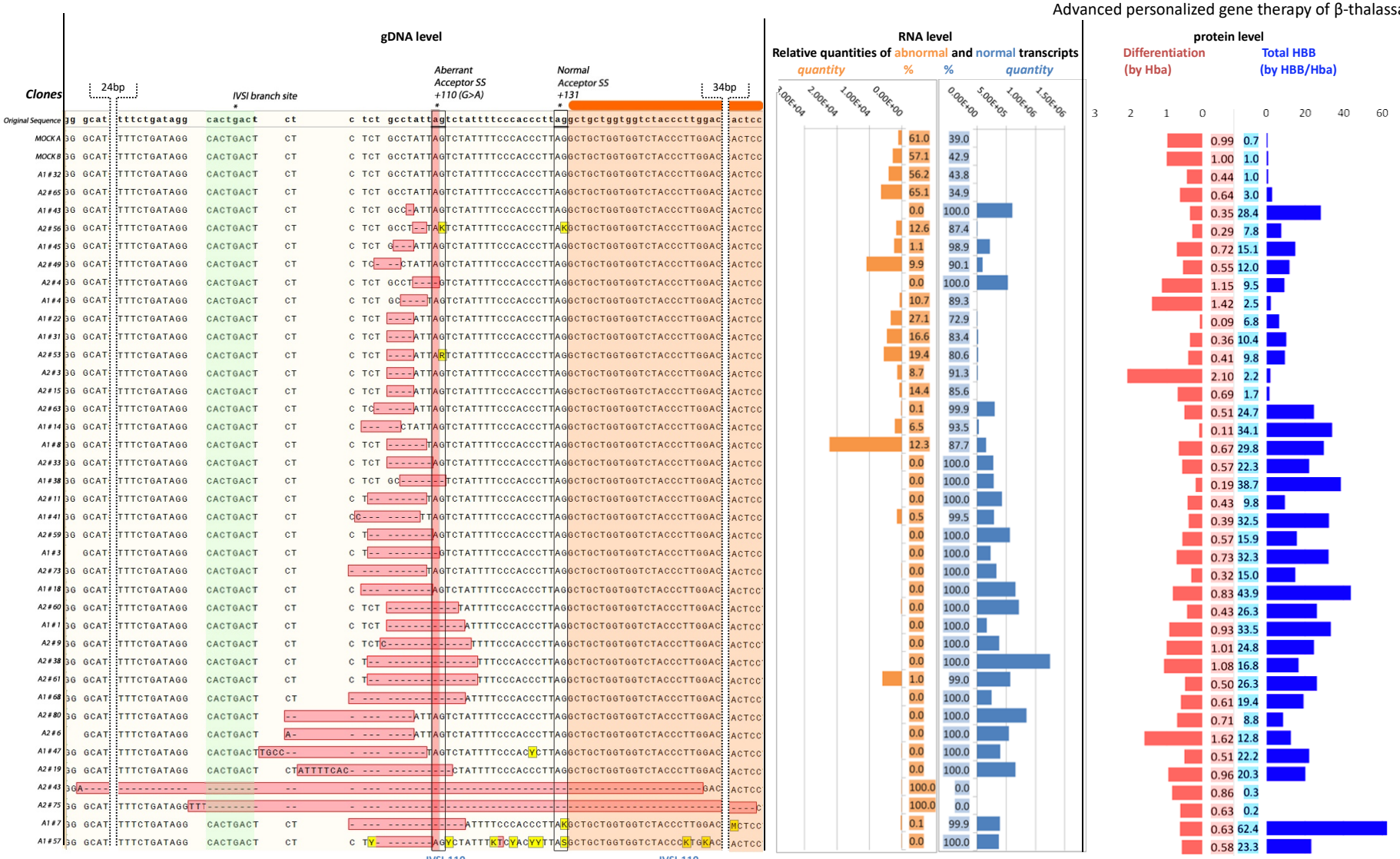


Figure 81: Correlation of specific modifications on the HBB$^{IVSI-110}$ region of MEL MA821 HBB$^{IVSI-110}$ VCN 1 clones with HBB expression at the transcriptional and translational level.
Alignment of gDNA sequences of differentiated MEL MA821 HBB$^{IVSI-110}$ VCN 1 clones on consensus HBB$^{IVSI-110(G>A)}$ sequence based primarily on the size of the INDEL and subsequently on the HBB$^{IVSI-110(G>A)}$ mutation (red vertical box). Clones treated with HBB TALEN pair R1/L1 or R1/L2 are referred as A1 or A2 # clones. MOCK A and B are non-edited negative control clones treated with pUC118 and A1 #32 , A1 #65 are non-edited clones treated with the corresponding TALEN pairs and both are used negative (–ve) controls. Orange boxes designate Exon II, green boxes the IVSI branch site and rectangle the abnormal and normal splice acceptor sites. Correlation of specific clones/INDEL with HBB mRNA and protein levels is indicated on the graphs. The mRNA levels graph exhibit quantities of abnormal and normal transcripts normalised relative to expression of total HBB per Hba (bars). Percentage (%) of **aberrantly** and **correctly** spliced HBB mRNA in the total HBB mRNA population is shown. The protein level graph shows the fold difference of **differentiation (via Hba)** and **normalised HBB/Hba levels** relative to –ve control (Mock B). Alignment of sequences was performed using SnapGene software.

3.3.6.1.3 Comparison of edited MEL HBB$^{IVSI-110}$ VCN 1 clones with MEL MA821 HBB^{Normal} cells

Protein expression comparison of the transgenic MEL HBB$^{IVSI-110}$ VCN 1 clone and MEL MA821 HBB^{Normal} VCN ~2 by immunoblots showed that on day-6 of erythroid differentiation approximately only ~4.2% of the differentiation-normalised HBB chains of the normal transgenic cell line, is produced in the mutated clonal cell line (Figure 26). Conversion of the 1 unit used to calculate the fold differences of differentiated HBB chains relative to the non-edited cells, MEL MA821 HBB$^{IVSI-110}$ VCN 1 to the 4.2 % of expression seen in the MA821 HBB^{Normal} VCN ~ 2, allowed measurements, which showed that the edited clones are expressing the HBB chains close to "normal" levels. In Figure 82 is illustrated the estimated percentage of each clone relative to the normal levels. Normal level is arbitrarily set as 100% and the error bars (+/- 22.18%) represents the fluctuation of the differentiation-normalised HBB chains levels seen on day-6 of differentiation of MEL MA821 HBB^{Normal} in three independent experiments (Figure 26). Expression levels within that range are considered to be fully functionally corrected. A total of 16 clones reached or even exceeded the normal levels from which 10 were edited by HBB TALEN pair R1/L1 (A1), 6 were between 50–70%, 5 between 30–40 and the rest were below the 12% background level of detection in the non-edited clone A2 #65. Based on these findings approximately, 50% of the editing events completely restored splicing and in the remaining 1/3 partial restoration, which should provide therapeutic benefit in β-thalassaemic patients.

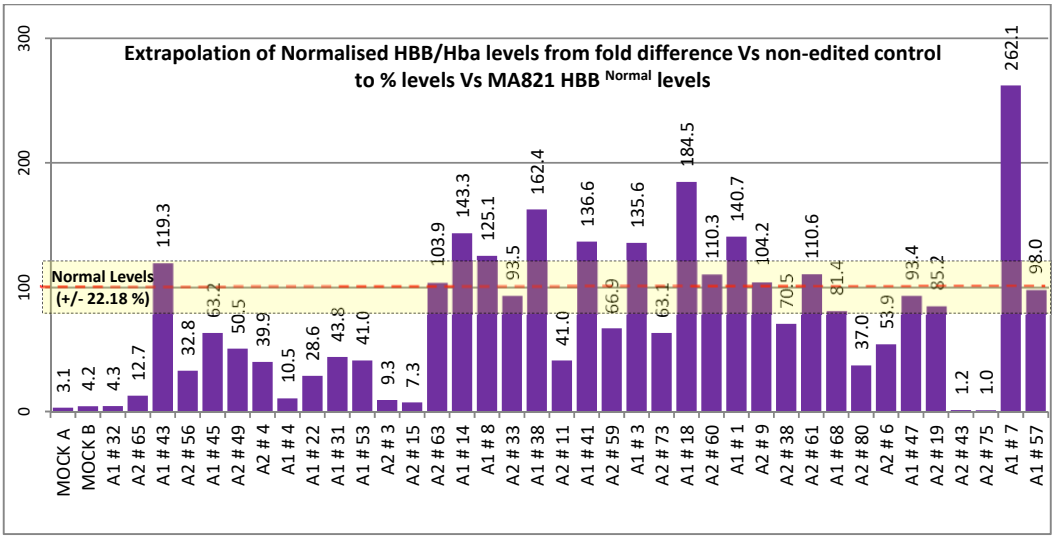


Figure 82: Estimated percentage of HBB chains in edited MEL MA821 HBB$^{IVSI-110}$ clone VCN 1 clones relative to those in MEL MA821 HBB^{Normal}.

Fold difference from non-edited MEL MA821 HBB$^{IVSI-110}$ clone VCN 1 on day-6 of erythroid differentiation corresponds to 4.2% percent of the expression observed in MEL MA821 HBB^{Normal} control. Normal level is set as 100% and the error bars (+/- 22.18%) represents the fluctuation of the degree of erythroid differentiation normalised HBB chain levels in MEL MA821 HBB^{Normal} in three independent experiments. Clones expressing levels within this range are considered to be fully functionally corrected.

Finally, an immunoblot analysis was conducted with clones differentiated in parallel (Figure 80) with all the transgenic MEL HBB$^{IVSI-110}$ and HBB^{Normal} cells as confirmation of the previous result and for a direct comparison of the restoration of splicing and level of differentiation at the level of HBB chain production compared. Immunoblots with protein extracts from day-6 of erythroid differentiation are shown in Figure 83. Quantification of the protein bands indicated that in general the level of differentiation was similar in all clones with the exception of A2 #73, which was lower by 80% compared to the MA821 HBB^{Normal} control. A large diversion in the degree of differentiation in some samples from the overall differentiation reduced the reliability and accuracy of the data that could be obtained from the comparison of the differentiation-normalised HBB chains with the rest of the samples, which can become misleading.

In particular in the case of A2 #73, low levels of differentiation probably lead to an over – estimation of normalised HBB chain levels. Edited clones A2 #63, A2 #61 and A2 #60 showed more than 50% of the expression observed in the negative control, whereas clone A2 #59 reached a normal level (94%) and A2 # 80 even exceeded the normal level (130%) (Figure 84). Even though the first differentiation (Figure 85), showed a lower level of HBB chains in these two (A2 #59 and A2 #80) clones, in general all the expected edited clones showed restoration of splicing of over 50% compared to normal samples at the level of HBB production, whereas the non-edited control remained as expected at 2.58%. In addition, the clones with large deletions namely A2 #43 and A2 #75, in both differentiation experiments did not express any HBB chains even though they were successfully differentiated (Figure 83 & Figure 85).

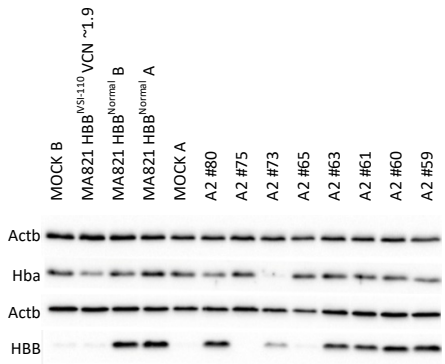


Figure 83: Immunoblot on protein extracts from the 2nd erythroid differentiation of genome-edited MEL MA821 HBB$^{IVSI-110}$ clones

Re-assessment of splicing restoration with a 2nd erythroid differentiation of edited clones shown on the 4th immunoblot (Figure 80) relative to the transgenic MEL HBB$^{IVSI-110}$ average VCN 1.9 and clone VCN 1, Mock A and B, and MA821 HBB^{Normal} average VCN 2 A and B. Whole cell extracts from 0.5–1x10^6 cells were probed with antibodies against HBB, Hba and Actb. Actb was used as a protein loading control. Quantification of protein bands for HBB and Hba chains was normalised for protein loading using Actb as a reference from protein extracts of MEL MA821 HBB$^{IVSI-110}$ edited clones on day-6 of differentiation. Band densities were measured using ImageLab.

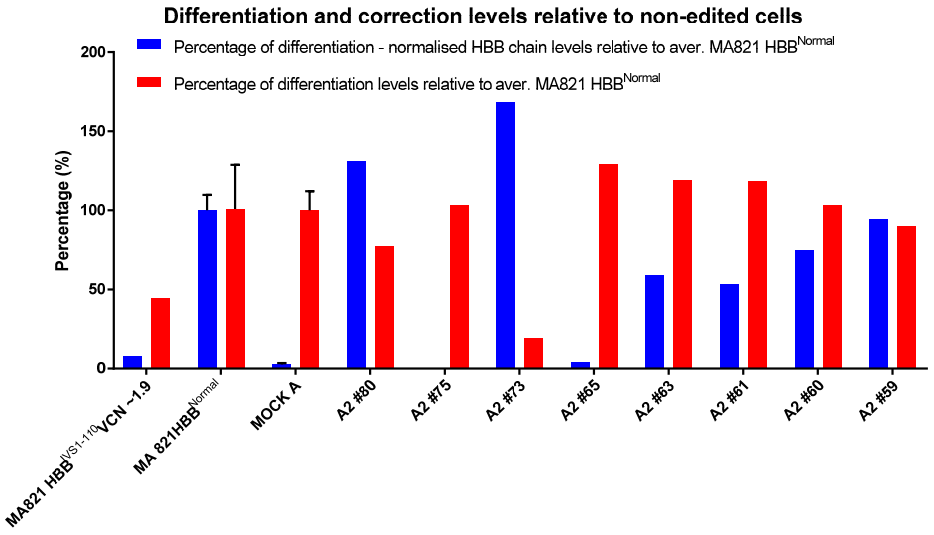


Figure 84: Correction of HBB chain levels from the 2nd erythroid differentiation of edited MEL MA821 HBB$^{IVSI-110}$ VCN 1 clones relative to MEL MA821 HBB^{Normal} average VCN 2.

Red bars indicate the percentage of differentiation (via Hba chains) and in blue bars the percentage restoration of HBB chain production (HBB / Hba chains) both relative to the MEL MA821 HBB^{Normal} positive controls. MEL MA821 HBB^{Normal} positive controls and non-edited samples were loaded twice on the immunoblot and the average percentage shown on the graph. MEL MA821 HBB^{Normal} positive control average VCN 2 and MEL MA821 HBB$^{IVSI-110}$ VCN 1 clone: differentiation level 100±27.8% and 100±12.07% differentiation-normalised HBB chains: 100±9.72% and 2.58±0.82%, respectively.

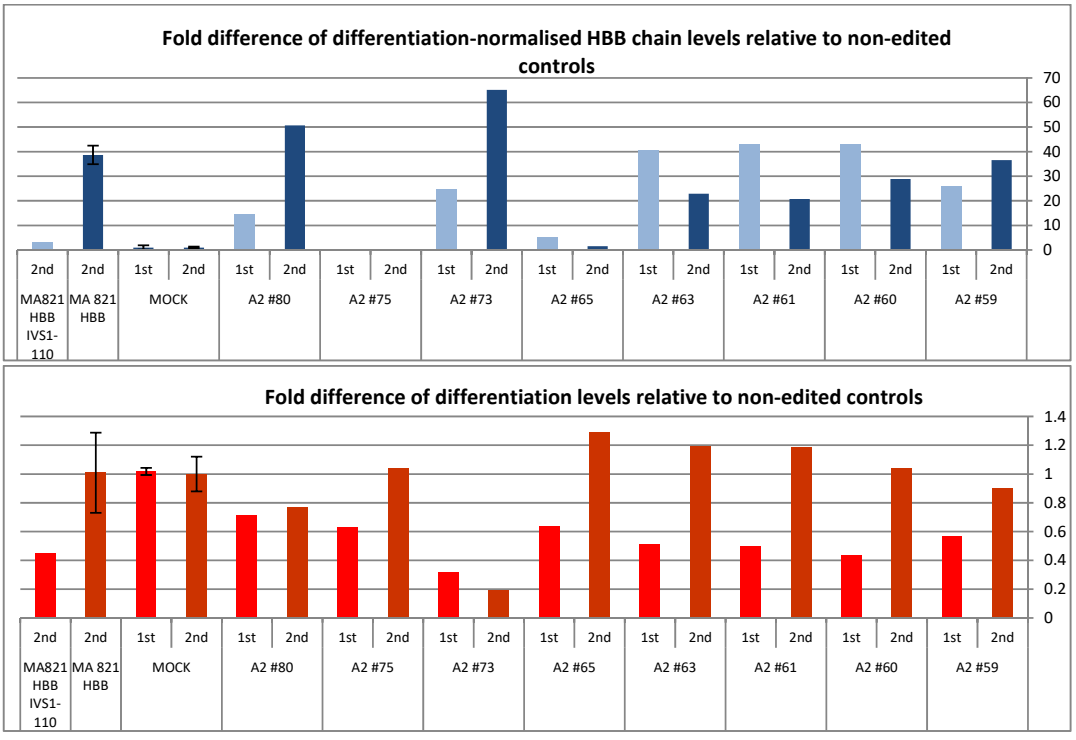


Figure 85: Levels of erythroid differentiation-normalised HBB chains and differentiation of genome-edited MEL MA821 HBB$^{IVSI-110}$ clones from two independent erythroid differentiation experiments.

Comparison of erythroid differentiation-normalised HBB chains levels (HBB / Hba chains) (blue bars) and differentiation level (via Hba chains) (red bars) as fold difference relative to non-edited MEL MA821 HBB$^{IVSI-110}$ VCN 1 negative (–ve) controls (Mock) in two differentiation experiments (1st light colour and 2nd darker colour) of the same clones. Non-edited –ve controls (Mocks) were performed in duplicated in both differentiations (1st and 2nd) (n=2). In the 2nd differentiation, MA821 HBB$^{IVSI-110}$ average VCN ~1.9 and MA821 HBB^{Normal} average VCN 2 (n=2) were added.

3.3.7 *Conclusions: MEL genome editing*

In this section, proof of principle was established in which functional correction of the *HBB*$^{IVSI-110(G>A)}$ mutation was achieved by targeted disruption of the abnormal SA site and the adjacent region through the NHEJ repair pathway in transgenic MEL HBB$^{IVSI-110}$ cells. Results showed restoration of pre-mRNA splicing leading to a corresponding increase in HBB chain production in cell populations edited with the most active HBB specific TALEN pairs, R1/L1 and R1/L2, with the HBB$^{IVSI-110}$ RGEN proving to be far less efficient. In addition, co-transfection of the designer nucleases with TREX2 did not noticeably alter the level of targeted disruption (as assessed by T7E1 assay/TIDE) or size of deletions (shown by TIDE) compared to samples edited without TREX2. Likewise, the efficiency of splicing restoration was not significantly improved, neither at the RNA nor protein level. TREX2 is a DNA end processing enzyme, which participates in the mechanism of DNA DSB repair and possesses a 3' to 5' exonuclease activity. Therefore inclusion of TREX2 in principle can provide more efficient/mutagenic INDELs when is combined with designer nucleases, which produce 3' end overhangs such as the homing endonucleases. TALEN-produced DSBs are mainly generated with 5' overhangs, whereas RGEN creates blunt ends [190,251]. Although DNA end processing enzymes such as TREX2 can increase mutagenesis rates when combined with specific designer nucleases, it is also associated with an increase in the level off-targeting effect [252]. HBB TALEN pairs have proven to be very efficient genome editing tools which allowed the establishment of the proof-of principle of our hypothesis without the help of TREX2. Since our final aim is the development of GT that could be applied into the clinic, including the TREX2 to further experiments is not considered essential. We believed the delivery of extra constructs, such as TREX2, in hCD34^{+} cells will probably have more adverse effects, such as increase of off-targeting and toxicity, than benefits.

In this section, the type of INDELs produced by the TALEN and RGEN genome editing approaches were characterised by Sanger sequencing of TOPO clones of PCR products spanning the IVSI-110 mutation in MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cell populations. Results show that TALENs mainly produced deletions as previously reported and that the pair with a larger spacer (13 bp) tends to produce deletions starting from 3–4 bp, whereas the TALEN pair with the shorter spacer (10 bp) gave deletions starting from 1–2 bp. This could be because cleavage by the FOKI monomers, due to the larger distance between the monomers, produces slightly larger overhangs, which are more permissive to losing more nucleotides than the shorter overhang before re-joining of the two ends.

The results obtained also show the level of splice restoration in clones with specific modification of the *HBB*$^{IVSI-110(G>A)}$ transgene. These data indicate that disruption of the region

5' to the mutation is sufficient to restore splicing, whilst leaving the mutation intact. This can be readily explained by the fact that the SA site consists of not only the AG sequence but also a PYT immediately 5', which is rich with pyrimidines mainly uracils and which is bound by the essential spliceosome factors U2AF1/2 and polypyrimidine-tract binding proteins (PTBs) [81,253]. Therefore alternation or deletion of the PYT as a result of the genome editing targeted here , especially with removal of T and C would reduce the efficiency of formation of the pre-splicing complex on the aberrant IVSI-110 SA site and hence force the spliceosome to use the normal SA site [81]. Such a scenario that graphically illustrates the sensitivity of the whole splicing procedure to an intact PYT is clone A1 #43, in which a deletion of a single T, 4 bp away from the aberrant IVSI-110 SA site is able to completely restore splicing to normal levels. Another factor that possibly contributes to the preference of the spliceosome to use the normal SA site instead of the aberrant IVSI-110 in edited clones, is the shortness of the distance of the normal branch site from both 3' splice sites. In a previous study it was shown that the abnormal SA site utilised the normal branch point, and therefore competes with the normal SA site resulting in reduced correct splicing [92]. In addition, restoration of correct splicing has also been achieved by inactivating the normal branch with specific antisense oligonucleotides [91]. This led to the activation of a cryptic branch site at nt 107 of intron 1, which is in close proximity with the aberrant SA site, that was mainly utilised by the normal SA site reversing the ratios of aberrant to normal spliced products from 9:1 to 1:5 in treated samples [91]. Hence, most likely a minimum distance is also required between the branch site and 3' splice site for the formation of a functional pre-spliceosome complex and the successful splicing of the intron. The distance of the branch site from the 3' aberrant and normal splice site is 14 bp and 35 bp, respectively. In this regard it is noteworthy that in cases where the aberrant SA site was left intact but the distance was reduced between 4–9 bp, namely edited clones A2 #63, A2 #33, A2 #11, A1 #41, A2 #59, A2 #73, A1 #18, A2 #80 A2 #6, A2 #47, complete restoration of correct splicing was observed, with exception of A1 #14 and A1 #8 where a small percentage (6% and 12.3 % respectively) of aberrantly spliced HBB mRNA was still detected. Furthermore, all clones that showed disruption of the IVSI-110 mutation, with exception of the two clones where the deletions were extended into Exon II, restored correct splicing. Both TALEN pairs R1/L1 and R1/L2 in general gave similar targeted disruption. However, TALEN pair R1/L1 tends to create deletions starting from 3–4 bp. Functional correction in clones was observed mainly in deletions starting from 5 bp and this also is in agreement with the observation that the majority of corrected clones are edited with TALEN pair R1/L1. Therefore, based on these observations it was decided experiments on patient-derived $CD34^+$ cells were performed with HBB TAL R1/L1.

3.3.8 *Genome editing of hCD34^{+} cells*

Three independent experiments were performed on hCD34^{+} cells mainly using the HBB$^{IVSI-110}$ TALEN R1/L1 pair at different amounts of plasmid DNA (40, 30 and 20 μg) and in parallel using the pUC118 nuclease-free as a negative control (30 μg). Electroporated cells without plasmid DNA were also included for the assessment of toxicity induced by the plasmid amount. However, evaluation of correction potential of designer genome editing nucleases was assessed relative to the pUC118 samples. DNA-free samples as expected recovered very well post-electroporation with low cell death and retained higher proliferation rates without showing signs of stress. In addition, during induced erythroid differentiation DNA-free controls displayed a different rate of differentiation in combination with low cell death, and in all experiments the ratios of HBB-like globins (*HBB* and *HBG*) were considerably higher than in pUC118 transfected negative control samples. Initially, at 48hr post-electroporation cells were harvested for the assessment of transfection efficiencies (percentage of GFP^{+} cells), cell death and percentage of genome edited cells by T7E1 assay. The remaining cells were then induced to differentiate down the myeloid/erythroid lineage. Based on the condition of the cultures, cells were harvested on day-7 for the assessment of restoration of pre-mRNA splicing (via multiplex RT-qPCR) and HBB chain production (via HPLC). A total of three independent experiments were performed with CD34^{+} cells from three patients (A, B and C).

Transfection efficiencies were similar between samples at 40–50 % in all the experiments and cell death was relatively low being less the 30% based on trypan blue positive cell counts (Figure 87). Transfection efficiencies of the 1st experiment was measured under fluorescence microscopy (three counts per sample), whereas in the 2nd and 3rd experiments by flow cytometry (Figure 86).

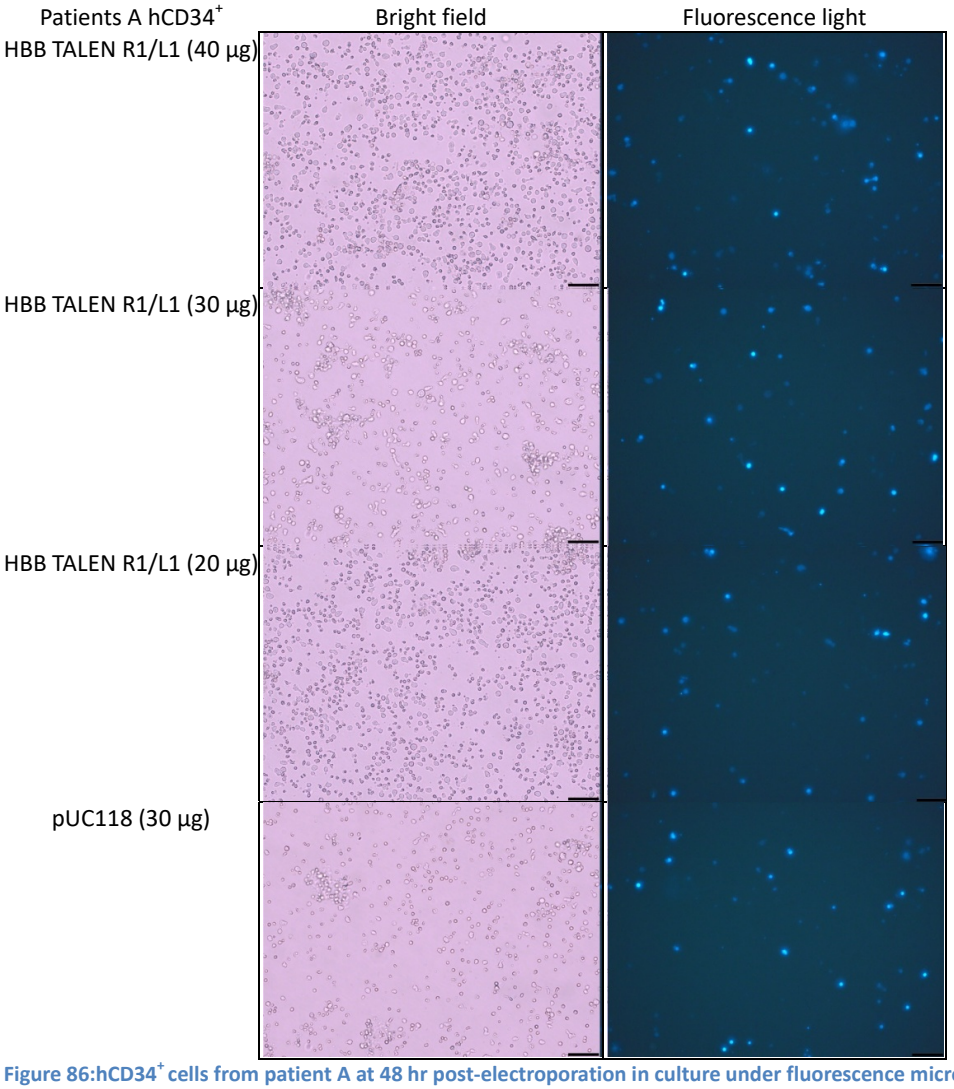


Figure 86:hCD34$^+$ cells from patient A at 48 hr post-electroporation in culture under fluorescence microscopy. Samples were transfected with HBB TALEN R1/L1 (40, 30 and 20 µg) and pUC118 (30 µg) as a negative control. All samples were co-transfected with 5 µg eGFP expressing construct. Images from fluorescence microscope observations were used to assess transfection efficiency (three counts/sample). Magnification 40 X, — Scale bar: 20 µm.

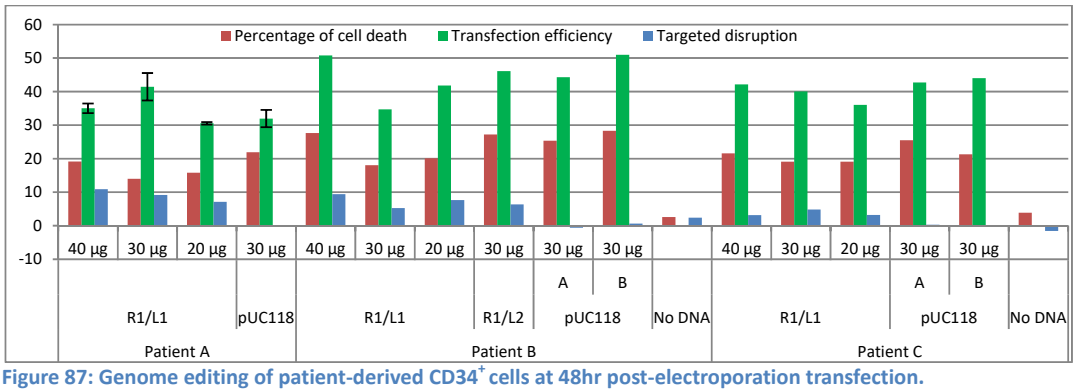


Figure 87: Genome editing of patient-derived CD34$^+$ cells at 48hr post-electroporation transfection. Transfection efficiency (green bars) was measured as the percentage of GFP$^+$ cells in patient A by fluorescence microscopy (three counts) and in samples from patients B and C by flow cytometry. Cells death (red bars) is shown as the percentage of trypan blue positive cells. Percentage of targeted disruption of the *HBB*$^{IVSI\text{-}110}$ gene was assessed by T7E1 assay (blue bars). Samples were treated mainly with HBB TALEN R1/L1, and once with R1/L2, at different quantities of plasmid DNA (40, 30, 20 µg). PUC118 was used as a nuclease-free negative control in which 30 µg had being used for the assessment of targeted disruption and electroporated cells without DNA (No DNA) as GFP negative control for the assessment of transfection efficiency via flow cytometry. Duplications within experiments are designated as A and B.

Even though a relatively high transfection efficiency was observed in electroporated CD34$^+$ cells, targeted disruption efficiencies measured by T7E1 assay (Figure 88) was low with the highest to be seen in the patient A HBB TALEN R1/L1 (40μg) sample at 12% (Figure 87).

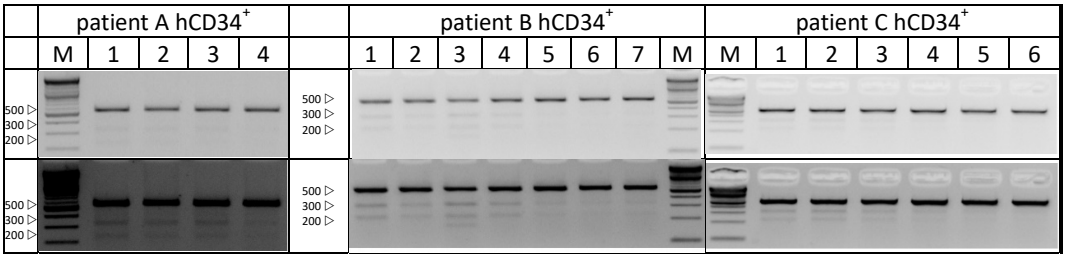


Figure 88: Targeted disruption *HBB$^{IVSI-110}$ gene* in patient-derived CD34$^+$ cells with HBB TALEN pairs.
Patient A: lane 1: (40μg), lane 2 (30μg), lane 3 (20μg) of HBB TALEN R1/L1 and lane 4 (30μg) of pUC118. Patient B: lane 1 (40μg), lane 2 (30μg), lane 3 (20μg) of HBB TALEN R1/L1, lane 4 (30μg) of HBB TALEN R1/L2, 5 and lane 6 (30 μg) pUC118. Lane 7 No DNA negative control. Patient C: lane 1 (40μg), lane 2 (30μg), lane 3 (20μg) of HBB TALEN R1/L1, lanes 4 and 5 (30 μg) pUC118, lane 6 No DNA negative control. M: 2log DNA ladder. Uncleaved band 500 bp, cleaved band 300bp and 200 bp. Bands were quantified using ImageJ software. Lower gels are saturated images of the top panel.

3.3.8.1 *Restoration of HBB pre-mRNA splicing and restoration of HBB chain production*

Genome-edited hCD34$^+$ cells were induced to differentiate as liquid cultures down the myeloid/erythroid lineage and samples collected for total RNA and protein extraction on day-7 post-differentiation. In experiments with cells from patients B and C, the negative controls, pUC118 and No DNA, differentiation was conducted in duplicate for a more accurate assessment of the levels in non-edited cells. Unfortunately, cultured-cells from patient B, edited with the HBB TALEN R1/L1 (30μg) constructs, became contaminated and therefore no functional data obtained from this sample. The risk of contamination with our optimised CD34$^+$ electroporation protocols is relatively high, since during the 48 hr expansion/recovery period post-electroporation, cells are cultured in P/S-free medium in order to increase cell survival after electroporation. The formation of pores on the cell membranes allows the antibiotics to enter the cell, which can induce cells death probably due to the formation of toxic metabolite intermediates.

3.3.8.1.1 Cytocentrifugations

Assessment of erythroid-differentiation on day-7 of differentiation with cytocentrifugation of cells on slides and stained with Giemsa, May-Grünwald and dianisidine. Dianisidine stains the cells expressing Hb. After cell counting of dianisidine positive cells from different fields of the stained slide we noted that after seven days of induced-differentiation, cultures have on average ~ 78±6.9% dianisidine-positive cells. In general all cultures have the similar percentages of Hb expressing cells with exception the negative controls that were

electroporated without plasmid DNA (No DNA –ve control). As mentioned above, these cells had different rates of amplification and differentiations and in general they looked healthier than cells that received plasmid DNA. Based on the dianisidine staining, the No DNA negative controls have in both experiments (B and C) on average ~10% higher positive cells than the pUC118 nucleases-free negative controls. Therefore, we believed that assessment of correction potential should be measured relative to the nucleases–free negative control (pUC118), in which cells were exposed to relatively similar conditions and stressed factors as edited samples. In addition samples treated with HBB TALENs pairs did not diverge from the pUC118 negative control, which indicates if there is any functional correction of the HBB$^{IVSI-110\ (G>A)}$ mutation due to genome editing that is expected to be marginal.

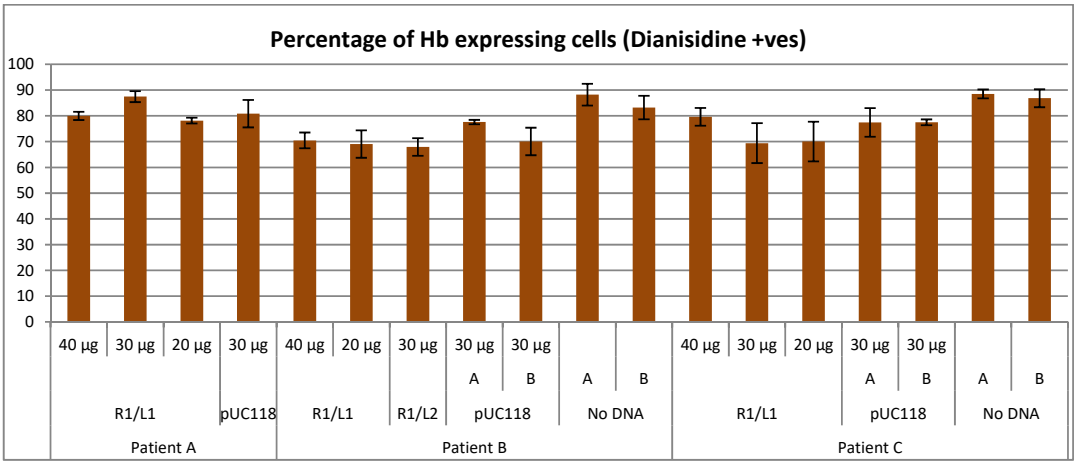


Figure 89: Differential counts of Dianisidine / May-Grünwald / Giemsa stained cytocentrifugation samples from day-7 erythroid differentiation cultures of electroporated hCD34$^+$ cells from three different IVSI-110 patients (A, B and C).

CD34$^+$ cells were electroporated with different amounts of total plasmid DNA (40, 30, and 20 µg) with designer nucleases constructs (HBB TAL R1/L1 and R1/L2). PUC118 nucleases-free and No-DNA samples serve the negative controls. Electroporated cells were induced differentiated for 6–7days in differentiation medium and harvested for further analysis, including the stage of differentiations. 5–10x10^4 cells were transferred on slides via cytocentrifugation and stained with Dianisidine / May-Grünwald-Giemsa staining. The percentage of Hb expressing cells (dianisidine positives) of each sample was assessed based on the cell counts from different fields on each slide, under the microscope. Note, due to the low targeted disruption efficiency of our designer nucleases in these set of experiments we have not proceed to the full phenotypic characterisation of the erythroid progenitors based on the are morphology.

3.3.8.1.2 Analysis of HBB mRNA

Analysis of the percentage of correctly and aberrantly spliced HBB mRNAs on day-7 of differentiation did not show an obvious change in favour of the normal mRNA. The ratios of normal to aberrant HBB mRNAs in cells treated with TALENs in all experiments remained close to that seen with the pUC118 nuclease-free negative control being 1.37 (n=1), average 0.9±0.06 (n=2) and 0.81±0.03 (n=2) in patient cells A , B and C, respectively (Figure 90).

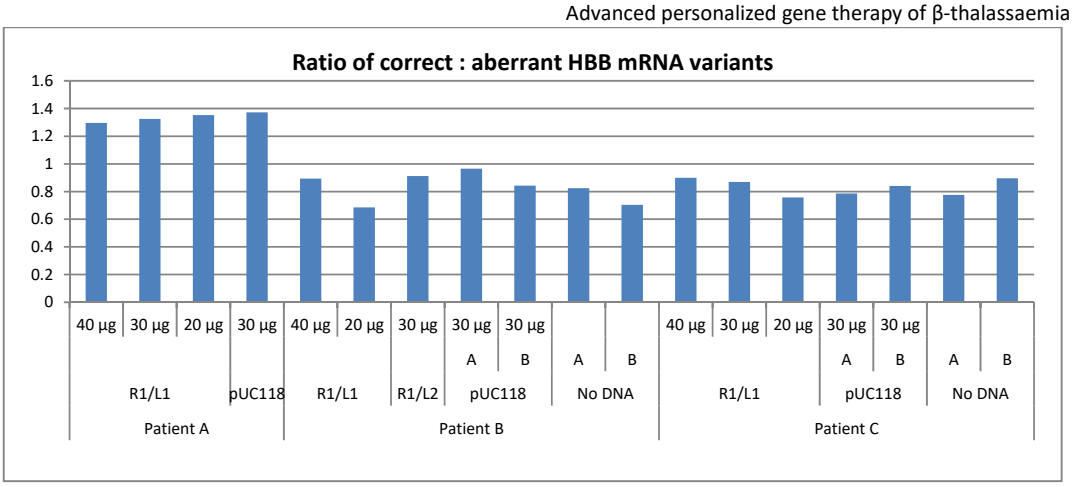


Figure 90: Assessment of the level of restoration of correct HBB mRNA splicing at day-7 of induced erythroid differentiation of genome edited patient-derived hCD34$^+$ cells (patients A, B and C).
CD34$^+$ cells were electroporated with different amounts of total plasmid DNA (40, 30, and 20 µg) with designer nucleases constructs (HBB TAL R1/L1 and R1/L2). PUC118 nucleases-free serve the negative controls. Electroporated cells were induced differentiated for 6–7days in differentiation medium and harvested for assessment of the level restoration of splicing at the RNA level, via multiplex RT-qPCR analysis. Absolute quantities of correct and aberrant HBB mRNA variants were measured using a plasmid-based SC. The graph shows the ratio of the normal : aberrant HBB mRNA in the total HBB mRNA population. Duplicates within experiments are designated as A and B.

3.3.8.1.3 Analysis for HBB chains

Restoration of HBB chain production was assessed by HPLC analysis. Protein extracts from day-7 of differentiated cells were analysed and globin chains were quantified. Globin chains were identified based on their elution time from the column. The level of restoration of splicing at the protein level can be illustrated as the percentage change of HBB/HBA and HBG-1/HBA chain ratios of genome edited samples compared to the nuclease free pUC118 negative control. It was also noticed that in differentiated CD34$^+$ cells, when HBB chains increase, HBG chains decrease, as expected. Figure 91 shows the HPLC data of edited cells from patient A treated with HBB TALEN R1/L1 at different amounts (40, 30 and 20 µg) and the pUC118 negative controls.

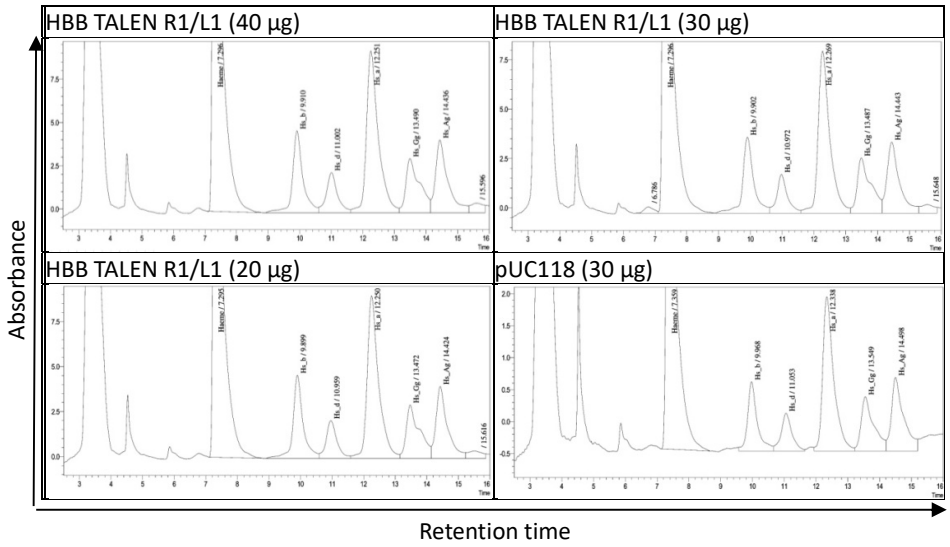


Figure 91: HPLC data quantifying globin chains from protein extracts from patient A genome edited CD34$^+$ cells on day-7 of differentiation.

Cells were edited by transfection with different amounts (40, 30 and 20 µg) of the HBB TALEN R1/L1 pair. pUC118 (30 µg) is used as nucleases-free negative control. Electroporated cells were induced for 7 days in differentiation medium and harvested for assessment of the level restoration of splicing at the protein level, via high-performance liquid chromatography (HPLC) analysis. Plots show peaks in the amount (absorbance) of each globin chain and is identified based on its retention time on the column (x axis), which is specific for each globin chain. Haeme: 7.2 min, Hs_b (HBB chains): 10 min, Hs d (HBD chains): 10.8 min, Hs a (HBA chains):12.3 min, Hs Gg (HBG1 chains):13.7 min and Hs Ag (HBG2 chains): 14.6. Ratios of the peak areas of HBB-like/HBA chains are used to compare the expression levels between samples. The above data correspond to the HPLC analysis of day-7 differentiated hCD34$^+$ treated cells from patient A.

In general it was noticed that in cultures from *HBB*$^{IVSI\text{-}110}$ patient hCD34$^+$ cells, the levels of HBB chains are higher than what would be expected to see in such thalassaemia major patients. The ratio of HBB:HBA chains in non-edited samples (pUC118 treated) is 0.4–0.5. HPLC analysis of blood from a healthy individual has a ratio of approximately 0.8–0.9. Based on bibliography, if patients produced approximately 10–20 % of HBB chains measured in healthy individuals, we would expect the ratio to be around 0.08–0.18. One possible explanation of the increased expression HBB chains is the medium and culture conditions used for the *in vitro* cultures. The already high background levels of HBB chains detected in non-edited cells makes the assessment of restoration of HBB synthesis following correction of splicing more difficult and the HPLC method less sensitive to detect minor changes. In all three experiments we observed a slight increase of the HBB:HBA ratio in edited samples (Figure 92), with the best results seen in patient A edited cells with an 11–14% change relative to the pUC118 negative control sample (Figure 93). Note, the pattern of HBB:HBA percentage change almost follows the pattern of targeted disruption and all edited samples also showed a decrease in the amount of HBG1 chains (Figure 93). Changes in HBB chain levels relative to non-edited cells are marginal. However, this was expected since a low degree of targeted disruption of the IVSI-110 mutation (≤11 %) was achieved (Figure 87).

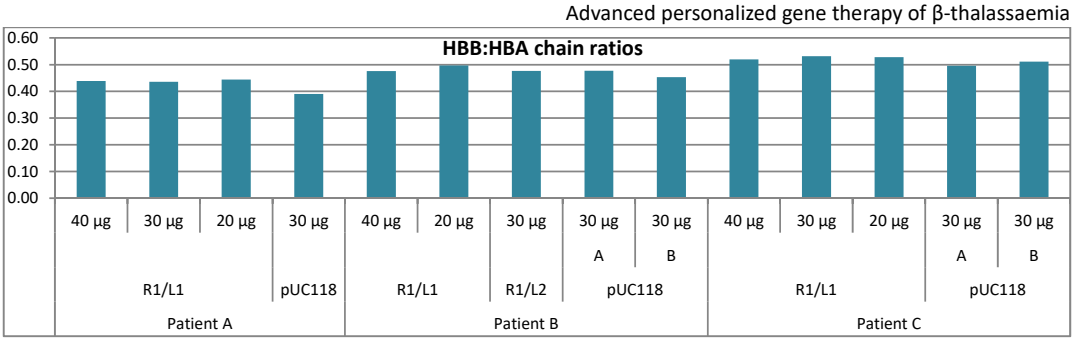


Figure 92: Ratios of HBB:HBA chain ratios in genome edited and non-edited patient-derived CD34$^+$ cells.
Ratios of HBB:HBA chains as measured by HPLC from edited patient derived hCD34 $^+$ cells following differentiation down the erythroid lineage for 7 days. Cells were edited by R1/L1 (40, 30, 20 µg) and R1/L2 (30 µg) whereas nuclease free pUC118 (30 µg) serve the negative control. Non - edited hCD34$^+$ (pUC118 treated) cells from three different patients (A, B and C) have a HBB:HBA ratio of ~0.4–0.5, which is higher (0.08–0.18) of what is expected in the blood of thalassaemia major patients of this type. Duplicates within experiments are designated as A and B.

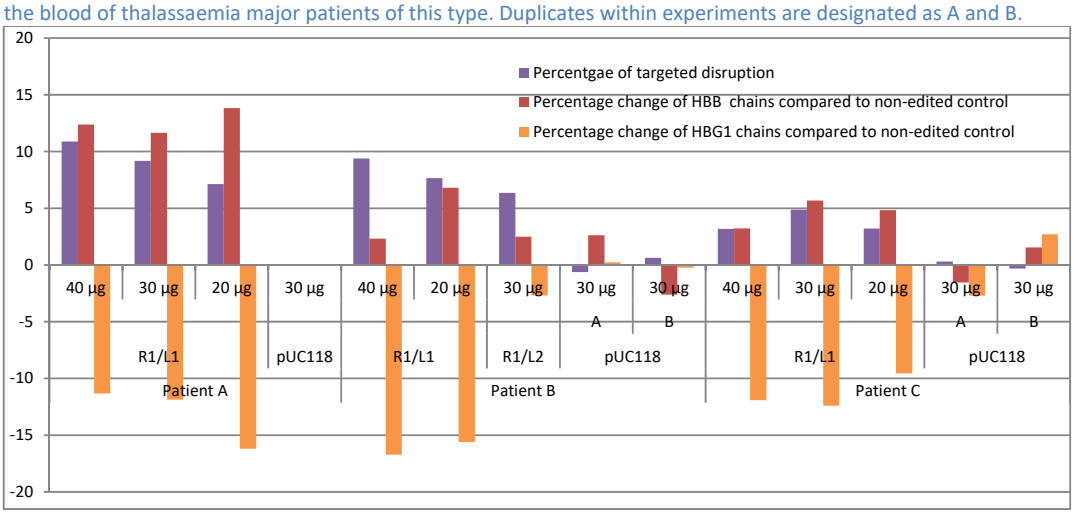


Figure 93: Percentage change of HBB:HBA and HBG1:HBA chain ratios in genome edited patient-derived CD34$^+$ cells relative to non-edited.
Percentage change in the HBB:HBA (red bars) and HBG:HBA (orange bars) chains in patient-derived CD34$^+$ cells edited by R1/L1 (40, 30, 20 µg) and R1/L2 (30 µg) relative to the nuclease free pUC118 (30 µg) negative control after induced erythroid differentiation for 7 days. The percentage of targeted disruption measured by T7E1 assay at 48hr post-editing and is shown in purple bars. Three independent experiments were performed using cells from three different patients (A, B and C). Duplicates within experiments are designated as A and B.

Fold increases in the HBB:HBA ratio compared to the nuclease free negative controls could be normalised on the percentage of genome edited cells in each population in order to get an estimation of expression levels in the edited cells only (Figure 94). Statistical analysis of the average fold difference of the HBB:HBA ratio indicated that edited cells with TALEN R1/L1 (20 µg) expressed 2.44 fold (*p* value = 0.0035; n=3) higher HBB chains than non-edited pUC118 control samples. The rest of the edited cells (HBB TALEN R1/L1, 40 and 30 µg) showed approximately a 1.8 fold increase but which was not statistical significant. However, putting together all the data from edited cells by HBB TALEN R1/L1 regardless the amount of plasmid DNA used in the transfection, statistical analysis showed that edited cells expressed double the amount of HBB chains compared to that seen in non-edited cells (*p* value = 0.0027). If this estimation is valid then, edited cells expressed an HBB:HBA ratio of ~0.8, which is at the level seen in healthy individuals.

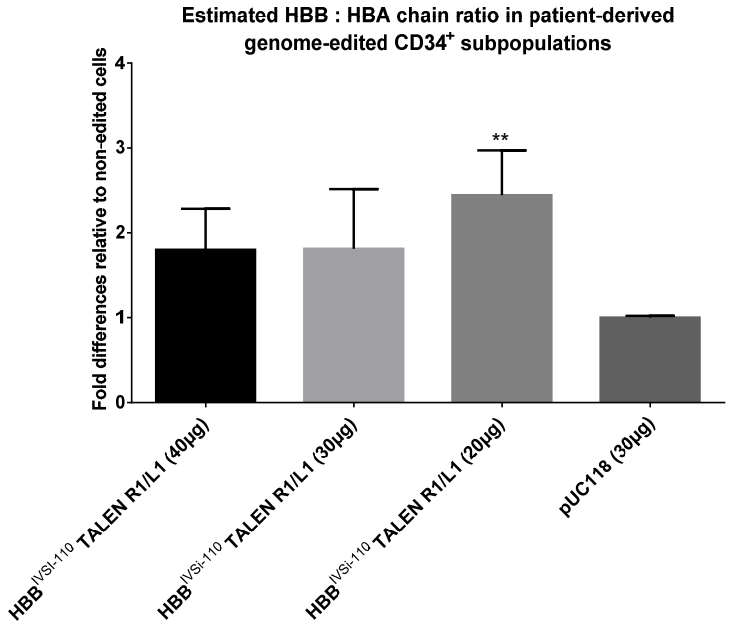


Figure 94: Estimated fold difference in HBB:HBA chain ratio in patient CD34$^+$ cells edited with HBB TAL R1/L1 40, 30 and 20 µg relative to nuclease free pUC118 negative control.

Statistical analysis showed edited cells by HBB TALEN R1/L1 20 µg has 2.44±0.52 fold more HBB chains than non-edited cells. Statistical analysis was performed by one-way ANOVA.

Statistical analysis

TALEN R1/L1 20 µg ** *p* value 0.0035

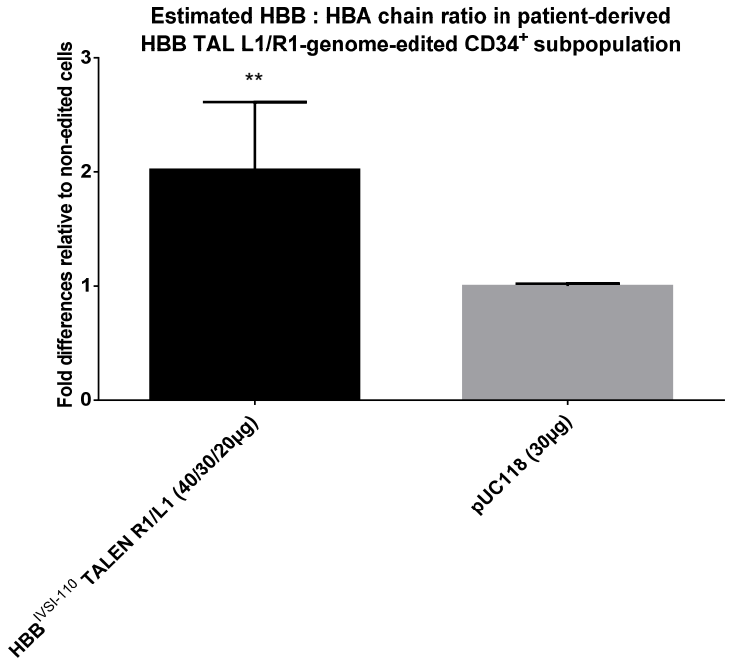


Figure 95: Estimated fold difference in HBB:HBA chain ratio in patient CD34$^+$ edited with HBB TAL R1/L1 relative to the nuclease free pUC118 negative control regardless the amount of delivered constructs (40, 30 and 20 µg).

Statistical analysis shows that cells edited with HBB TALEN R1/L1 have ~1±0.59 fold more HBB chains than non-edited cells. Statistical analysis was performed by one-way ANOVA.

Statistical analysis

TALEN R1/L1 (40/30/20 µg) ** *P* value 0.0025

3.3.9 *Conclusions: Genome editing with patient-derived CD34$^+$cells*

In this section application of the designer genome editing nucleases on patient-derived CD34$^+$ cells was assessed for their ability to restore normal splicing of the HBB$^{IVSI-110}$ pre-mRNA and HBB chain production. Before proceeding to genome editing experiments of the patient hCD34$^+$ cells, a series of experiments was performed to optimise the transfection protocol, using a conventional electroporation unit (Gene Pulser, Bio-Rad) and plasmid DNA. Even though, relatively high transfection efficiencies (up to 40–50 % GFP$^+$ cells) was achieved with low cell death, the percentage of targeted disruption remained low (>12%). Analysis of restoration of mutant HBB pre-mRNA splicing at day-7 post-erythroid differentiation of edited cells, showed that the percentage of correct and aberrant HBB mRNA remained the same with the non-edited controls. However, analysis at a protein level showed an increase in the HBB:HBA chain ratio, albeit to a marginal degree, relative to the non-edited controls. In general, the observed low levels of correct splicing restoration were not a surprise, since the percentage of edited cells in the population was low. Moreover, RNA data from the last day of differentiation (day-7) are not reliable and sensitive enough to detect minor changes between the two HBB variants. Cells for RNA analysis should be collected at an earlier time point (for example days 2–3), where transcription is still active and RNA levels are high. However, the low cell number during differentiation, in combination with the requirement for large numbers of cells for RNA and protein extraction, allowed the collection of cells only at one time point. Therefore it was decided to collects cells only at the last day of differentiation, which is ideal for protein extraction and globin chain quantification. Taking this into account the best indication of the restoration of splicing are those from protein analysis; that is, HBB chain synthesis measurements and not RNA. Estimation of the fold increase in the HBB:HBA chain ratio seen in the edited subpopulations indicated a doubling in the amount of HBB chains compared to non-edited cells. Based on the fact that non-edited patient CD34$^+$ cells have a relatively high background in their HBB:HBA ratio of 0.4–0.5, whereas the globin chain ratios from the blood of a healthy individual is between 0.8–0.9, the estimation in edited cells shows that they should have reached relatively normal levels. In general the levels of HBB-like and HBA-like chains is balanced in a 1:1 ratio at any developmental stage in normal cells, through transcriptional, post-transcriptional and post-translational control [254,255]. Therefore an increase beyond 0.9 in the ratio in edited cells will negatively be regulated to maintain a HBB:HBA normal balance.

In summary, the data generated in this work has provided a clear proof for the starting hypothesis that restoration of correct splicing by disrupting the aberrant IVSI-110 SA site and the adjacent poly-pyrimidine track region is valid in patients' hCD34$^+$ cells and that the

observed marginal restoration of splicing in $CD34^+$ cells is mainly due to technical limitations. As mentioned above, delivery of the designer TALEN and RGEN nucleases as plasmid DNA using conventional electroporation is not ideal for genome editing of $CD34^+$ cells mainly due to the high toxicity induced by the plasmid DNA. In general, for genome editing of $CD34^+$, research groups prefer to deliver the nucleases as mRNA, which provides a faster and short lived spike of expression but avoids prolonged exposure of the nucleases to the cells, thereby minimising off-target potential and genotoxicity. In addition, the use of new generation of electroporators with optimised protocols for electroporation of $CD34^+$ cells are now commercially available, such as Nucleofectors (*Lonza*) or the Neon electroporator (*Invitrogen*), which have been shown to provide low cell death, high transfection efficiencies and high targeted disruption. In addition, as mentioned above the high background of HBB chains in patient $CD34^+$ cells reduced the sensitivity of the HPLC analytical method for globin polypeptides. Therefore the percentage of edited cells needs to be increased in order to have a more profound difference in restoration of splicing and increase in HBB chain synthesis compared to non-edited cells. To this end, the TALEN constructs used in this study have been inserted into new plasmid vector backbones (pPIX_K_TALEN_FOK1) to allow their transcription to mRNAs for subsequent delivery to patient $hCD34^+$ cells using the Lonza nucleofection system. The new TALEN constructs, pPIX_K_TALEN_FOK1, employ a T7 promoter to drive *in vitro* transcription and a Kozak sequence for efficient translation once introduced into mammalian cells. It is anticipated that this approach will increase the percentage of genome edited $CD34^+$ cells and allow the assessment of the therapeutic potential of the overall therapeutic in patient $hCD34^+$ cells.

4 General Discussion

Among the haemoglobinopathies, β-thalassaemia is a common monogenic disease caused by more than 400 mutations with phenotypes categorised in three main groups, major, intermediate and minor. Thalassaemia major is the most severe form of β-thalassaemia, and patients require regular red blood cell transfusions from an early age in order to survive. The most severe forms under the large category of the milder β-thalassaemia intermediate, in which there is some production of functional haemoglobin, may also require occasional red blood cell transfusion. Finally, cases of thalassaemia minor are generally clinically asymptomatic with moderate anaemia in relatively few instances [57,61].

Research groups trying to develop GT for β-thalassaemia to date have focused on a gene-addition approach using LVs as delivery vehicle, with a number of clinical trials now underway [115,131,256] (online update on https://clinicaltrials.gov/ct2/home). More recently, genome editing has increasingly also been explored as a possible route of GT, which exploits the discovery of easily designed, programmable nucleases such as TALENs and in particular RGENs [39]. Although many years of optimisation of LVs have finally led to a satisfactory level of correction for several groups to enter clinical trials, efforts for further optimisation of vectors are still ongoing, such as the incorporation of small insulators, barrier elements and erythroid-specific enhancers, in order to achieve higher expression of the therapeutic transgene with the objective of minimising requirement of vector integrations and increasing safety [114,257].

In cases of β-thalassaemia where the underlying mutation or existing modifiers interfere with the effectiveness of gene addition, application of more personalised GT approaches would provide a significant therapeutic advantage. The $HBB^{IVSI-110}$ mutation is one such mutation, the presence of which reduces HBB production from normal or vector-encoded alleles and thus reduces the efficiency of therapy by gene addition [39,222,225].

The general aim of my PhD project was the development of advanced personalised GT for β-thalassaemia caused by the $HBB^{IVSI-110}$ mutation, which is frequently found in certain population groups and causes missplicing of intron 1. The two developmental approaches taken were optimisation of the efficiency of already pre-existing *HBB* LVs, such as MA821_HBB ("GLOBE"), or genome editing using specific designer nucleases based on TALENs and RGENs. Development of personalised GT of the $HBB^{IVSI-110}$ mutation would probably be of commercial and clinical interest, especially for countries or regions with high carrier frequencies, such as in Cyprus and Sardinia [54]. In addition, with the development of GT targeting the $HBB^{IVSI-110}$ mutation, similar strategies could in principle be applied to over 180 other molecular gain-of-

function mutations of regulatory regions outside open reading frames (our own unpublished survey), to target human disease beyond just β-thalassaemia.

4.1 Development of model cell lines and quantification methods

4.1.1 *Development of HBB$^{IVSI-110}$ model cell line*

One of the main requirements for the development of personalised GT is the availability of a model system that accurately mimics the disease condition at a genetic level. In particular cell lines inexpensively allow initial testing of hypotheses before moving to primary patient cells and more resource-intensive technologies for detecting and measuring the outcome of the applied therapy. Therefore MEL cell lines harbouring *HBB$^{IVSI-110}$* and *HBB^{Normal}* transgenes were generated by transduction with the MA821_HBB^{Normal} LV and its mutant derivative MA821_HBB$^{IVSI-110}$ LV. Mutant and normal transgenic MEL cell pools were selected for a similar average VCN, ~1.9 and ~2 respectively, in order to ensure relatively equal levels of expression. In addition, clonal lines of transgenic MEL MA821_HBB$^{IVSI-110}$ cells with VCN=1, were also isolated to better understand and assess the potential therapeutic advantage of the GT approaches developed here. Development of transgenic MEL cell lines using the MA821_HBB^{Normal} and HBB$^{IVSI-110}$ LVs have many benefits that expedited the work undertaken in this project. First, the mini-βLCR plus promoter of the MA821 HBB LV provided erythroid-specific expression of the *HBB* transgene, following induced erythroid differentiation of the MEL cells. Second, using the erythroid MEL cell line allows the simultaneous co-expression of endogenous *Hba* chains following erythroid differentiation, which can be utilised for the normalisation of HBB expression. While the vector-encoded HBB does not represent the whole *HBB* locus, it contains all sequences conceivably of relevance to our GT approaches, including the mini-βLCR in the MA821 *HBB* LV, which acts as a strong erythroid specific enhancer, generating and maintaining an open chromatin structure required to protect against silencing and long-term expression [115,132]. One of the most essential characteristics of these cell lines is that vector-derived HBB expression could easily be distinguished from the murine equivalent, at both a transcriptional and translational level, which would have been more difficult to determine if a human erythroid cell line (e.g., HEL) had been used.

4.1.2 *Development of a multiplex RT-qPCR method for HBB mRNA variant quantification*

One of our major concerns was the accurate and sensitive quantification of the two HBB variant transcripts, normal and aberrant, produced by expression of the *HBB*$^{IVSI-110}$ or *HBB^{Normal}* transgenes, in which the only distinguishable difference between the two variants are 19 nt

from intron 1, which is abnormally spliced into the mature non-functional mRNA. Previous studies used semi-quantitative methods, such as quantification of the two transcripts based on band densities upon separation on agarose or polyacrylamide gels. These methods, however, have low sensitivity, accuracy and reproducibility, all essential properties necessary for the evaluation and establishment of novel gene-therapy approaches. Therefore a new multiplex RT-qPCR, which allows the detection of both variants in the same reaction using variant-specific probes, was developed. Quantification of the two variants was measured by comparison to results from a plasmid standard curve in which the construct was designed to hold the distinguishable cDNA regions of both variants. According to the quantities loaded for the standard curve, the method employed here reached a high level of sensitivity, down to 40 template molecules per reaction. In general, this method was used for the assessment of changes in the ratios of the HBB mRNA variants. In addition, it was also possible to combine the absolute quantities of the two variants, obtained by multiplex RT-qPCR, with the relative quantification of total HBB expression normalised to erythroid differentiation (based on the ΔΔCt method). This method was specifically used for the correlation of INDELs with restoration of splicing in the genome-edited MEL MA821 $HBB^{IVSI-110}$ VCN 1 cell clone.

4.1.3 *Functional characterisation of transgenic MEL cell lines relative to patient-derived hCD34^{+}*

The newly developed humanised transgenic cell pool MEL MA821 $HBB^{IVSI-110}$ with average VCN 1.9 and the corresponding clonal cell line with VCN 1 were differentiated in parallel with transgenic cells containing MA821 HBB^{Normal} with average VCN ~2, and analysed at both transcriptional and translational levels. The results obtained showed that in both $HBB^{IVSI-110}$ transgenic cells (pools and clone); the aberrant mRNA constitutes ~ 60% of total HBB mRNA. Similar expression patterns were also detected in differentiated patient-derived $CD34^{+}$ cultures. The data obtained here are in agreement with a recent study, which showed that aberrant mRNA from *HBB*$^{IVSI-110}$-patient-derived erythroid progenitor cells is relatively stable and constitutes ~49% of total HBB mRNA [222]. Other groups have shown that in HeLa and monkey kidney COS cells expressing an *HBB*$^{IVSI-110}$ transgene, only approximately 10–20% of the pre-mRNA is correctly spliced and the rest is aberrant [82,239]. A more recent study in the humanised *HBB*$^{IVSI-110}$ mouse model system showed a different pattern of the aberrant to normal HBB mRNA ratio in a tissue-specific manner, where in spleen and bone marrow the predominant variant was the aberrant mRNA, while in PB it was the normal equivalent. Similar ratios were observed in patiental PB [90]. In most of the liquid cultures of erythroid cells from *HBB*$^{IVSI-110}$-patient-derived mononuclear cells, correctly spliced mRNA constituted ~66% of the total HBB mRNA. However, there were also cases where aberrant mRNA was not detected [94].

This difference in the correct:aberrant HBB mRNA ratios in different cell lines or tissues may be due to instability stemming from degradation via the NMD pathway of the aberrant mRNA. Of note, the relative mRNA quantities measured in the above studies were based on semi-quantitative methods, especially electrophoresis band density quantification of the two variants after RT-PCR, which is characterised by limited sensitivity and reproducibility. Most importantly, the results of this thesis showed that synthesis of HBB chains in transgenic $HBB^{IVSI-110}$ MEL cells relative to the normal counterpart, reflects the protein pattern seen in $HBB^{IVSI-110}$ thalassaemic patients, who achieve only approximately 10% of the HBB chain synthesis seen in healthy individuals [90]. Thus, the $HBB^{IVSI-110}$ MEL cell lines developed here represent a reliable model system, which recapitulates, at least at the translation level, HBB expression in PB of $HBB^{IVSI-110}$ thalassaemic patients.

4.1.4 *Summary of Conclusions: Development of model cell lines and quantification methods*

- We developed a cost-effective and robust multiplex RT-qPCR method for the accurate and sensitive quantification (down to 40 molecules) of the aberrant and normal HBB mRNA variants.

- Development and characterisation of humanised $HBB^{IVSI-110}$ and HBB^{Normal} transgenic MEL cell lines indicated at the transcriptional and translational level that they constitute reliable model cell lines for the investigation of the $HBB^{IVSI-110(G>A)}$ mutation

    - In multiplex RT-qPCR assays, $HBB^{IVSI-110}$ cell lines showed a ratio of splice variants (~54.5% is aberrantly spliced) similar to that observed for patient-derived $HBB^{IVSI-110}$ pre-mRNA (~60% is aberrantly HBB spliced) on day 7 of $CD34^{+}$ differentiation.

    - In immunoblots, steady-state levels of HBB chains in transgenic $HBB^{IVSI-110}$-transgenic MEL cells lines relative to the normal counterpart reflect the protein pattern seen in $HBB^{IVSI-110}$ thalassaemic patients, who achieve only approximately 10% of the HBB chain synthesis seen in healthy individuals.

4.2 Personalised gene therapy by gene addition

4.2.1 *MEL MA821 HBB$^{IVSI-110}$ transgenic cell line studies*

Based on previously published results [222,223] it was decided to verify the proposed "traffic jam" hypothesis for the optimisation of the therapeutic potential of HBB LVs in a mutation specific manner. In the study presented here the aim was to target/knock-down the aberrant mRNA with the expression of four shRNAs with overlapping target sequences (shIVSI-110 Up, Mid and Mid2 and Down) of the misspliced HBB IVSI-110 mRNA region. Mid2 is a modified version of Mid, in which the TTTT repeat in the antisense sequence of the shRNA was altered to TTGT, in order to avoid potential impairment of expression from the pLKO.I U6 cassette, as RNA pol3 could possibly recognise the TTTT as a termination signal [242,243]. On the one hand, assessment of silencing potential at the transcriptional level in the MEL MA821 HBB$^{IVSI-110}$ VCN 1 clonal cell line showed that only the shIVSI-110_Down was able to significantly knock-down the aberrant mRNA, whereas for the rest the ratios remained unchanged. On the other hand, at the protein level both shIVSI-110_Mid versions showed a significant increase in the translation of HBB chains, with slightly higher levels seen with shIVSI-110 Mid2. ShIVSI-110 Down displayed a marginal but not significant increase in HBB chains relative to the untransduced control. These preliminary data constitute the first indication that translation of the functional HBB mRNA (endogenously or vector-derived) can be significantly increased by specifically targeting the destruction of the aberrant HBB IVSI-110 mRNA. The data obtained suggest that only the shRNA that specifically targets the aberrant mRNA is able to increase the translation rate of the correctly spliced mRNA, whereas the others in which part of the sequence was complementary to the normal mRNA were unable to do so. Nevertheless, the results obtained strongly indicate that shIVSI-110 MID is responsible for an increase in the synthesis of HBB chains, although the precise mechanism responsible for this effect is still not clear, since RNA analysis did not show knock-down of the aberrant mRNA. A possible alternative mechanism of silencing is via translational repression and/or mRNA destabilisation [150]. This pathway of silencing is used by miRNAs, when there is an imperfect binding to the target mRNA sequence, usually in the 3'UTR region [258]. Although, our shIVSI-110-Mid RNAs are designed to perfectly match the aberrant mRNA, it is still a possible theory that is worth investigating. Noteworthy is the fact that the increased synthesis of human HBB chains were derived from a proportion of correctly spliced HBB mRNA expressed by the *HBB*$^{IVSI-110}$ transgene.

4.2.2 *Patient$^{IVSI-110}$-derived CD34^{+} cell studies*

Since the final aim of this approach is the increase of the therapeutic potential of *HBB*-like vectors, such MA821 HBB^{Normal} ("GLOBE"), in a mutation specific manner, we validated the

efficiency of all of our shIVSI-110 RNA constructs in IVSI-110 hCD34^{+} patient-derived cells. Subsequently, we used the most efficient and promising shRNA in combination with the MA821 HBB LV.

In general, hCD34^{+} were efficiently transduced with the pLKO.I shRNAs vectors resulting in very high VCNs (VCN ~25.5), whereas transductions with the MA821 HBB LV were relatively lower (VCN~2.7). Increased cell death was observed mainly in samples transduced with pLKO.I shRNAs LVs with obvious changes in the morphology of the cultures on the first day of induced erythroid differentiation. Increased cell death could be the result of increased genome instability and disruption of vital genes due to the high VCN [245], which in combination with the stress caused by the sudden change of culture medium, made them more permissive to apoptosis. Protein analysis of the HBB/HBA chain ratios in single shIVSI-110 transduced cells, was in agreement with the preliminary data obtained in the MEL transgenic cells, in which the shIVSI-110_Mid2 showed a significant increase of HBB chain production followed by shIVSI-110 Mid and at lower levels by shIVSI-110 Down. Although cells transduced with the shIVSI-110_Mid2 construct exhibited slightly higher HBB chains than those with shIVSI-110_Mid, it was decided to proceed with the latter for the double transduction experiment with MA821 HBB^{Normal} LV. The slight increase in HBB chains with shIVI-110 Mid2 relative to Mid, might be due to the higher expression of the shRNAs by RNA pol3 owing to the TTGT modification. However, as the final aim was co-expression of the shIVSI-110 RNA and the therapeutic *HBB* transgene from the same transcript following the strategy of Samakoglu and colleagues [158], wherein expression will be driven from the RNA pol2 *HBB* promoter, the presence of TTTT would not affect the expression level of the shIVSI-110 Mid.

Similarly, with the inclusion of MA821 HBB LV to the transduction experiments, we have shown an increase of HBB mRNA translation efficiency in cells expressing the shIVSI-110 Mid relative to the negative controls. The most surprising data were obtained from the comparison of MA821 HBB transduced cells (VCN ~2.7) and those with just shIVSI-110 Mid, in which the HBB:HBA chain ratio of latter even exceed the levels of the former.

In addition, differential counting of patient-derived CD34^{+} cells, showed a significant increase of the percentage of orthochromatic/reticulocyte cells, in single transduction with shIVSI-110 Up , shIVSI-110-Mid and especially shIVSI-110-Mid2 LVs, as also and in double transductions with shIVSI-110 Mid/MA821 HBB LVs. This feature constitutes an indication of the correction of the thalassaemic phenotype, which blocks normal erythroid differentiation, as less cells mature successfully to the orthochromatophilic and increased cell death is detected at the polychromatophilic stage [115]. Data from differential counting was in agreement with findings of protein expression levels which together further re-enforce the starting hypothesis. Moreover,

a single experiment performed with all the pLKO.I shIVSI-110 LVs on $CD34^+$cells from a healthy donor demonstrated a reduction in the HBB:HBA chain ratio only in the case of shIVSI-110 Down, whereas the rest of the shIVSI-110 constructs, and especially Mid and Mid2, exceed the normal levels. These data indicated that the shIVSI-110 RNAs tested, with exception of the shIVSI-110 Down, do not cause silencing of the correctly spliced mRNA, at least not to a degree that could be detected as a decrease of HBB chains below normal levels. However, as mentioned above, this constitutes data from a single experiment, and therefore more repeats should be conducted in order to allow reliable conclusions.

A recent study showed that the levels of HbA were increased in a VCN-depended manner in β^0/β^0 patient-derived erythroid progenitors cells transduced with the T9W LV, whereas in the case of β^+/β^+ and β^+/β^0 patient-derived cells, the response was slight or not at all to the LV, even though high levels of vector-derived HBB transcripts and HBA chain aggregates were detected [222]. Transduction of β^+thalassaemic cells with the AnkT9W LV restored the synthesis of HbA in a vector depended manner with a decrease in HBA aggregates, whereas in β^0/β^0 cells no major differences were observed between cells transduced with T9W and AnkT9W [222]. Based on this data the authors proposed the "traffic jam" hypothesis, in which they suggested that the relatively stable aberrant HBB mRNA blocks the translation of the normal HBB mRNA, which could be overcome with the addition of an ankyrin insulator element. They suggested addition of the ankyrin insulator allows higher expression of the therapeutic transgene at earlier stages of erythroid differentiation allowing the HBB transcript to efficiently compete with the endogenous aberrant mRNA for translational factors [222]. The data obtained in this project is in support of this hypothesis, where it was attempted to remove the competition between the two HBB variants by shRNA-targeting the aberrant HBB mRNAs with the expression of specific shIVSI-110 Mid/Mid2 RNAs at an early stage of differentiation, as they are driven from the constitutively expressed U6 promoter. Even though aberrant IVSI-110 mRNA knock-down was not detected in the transgenic MEL MA821 $HBB^{IVSI-110}$ clonal cell line, protein analysis in both MEL and patient-derived $CD34^+$ cells, clearly indicated a significant increase of the translation of the endogenously and/or vector-derived normal HBB transcripts.

Thus according to the data presented here the targeting of aberrant HBB mRNAs via shRNAs could constitute an effective element for inclusion in HBB LVs, which will increase substantially their therapeutic potential in patients with misspliced mutations. This would result in lower LV doses to reach a therapeutic threshold during clinical application thus increasing the GT margin of safety.

4.2.3 *Summary of Conclusions: Gene addition approach*

- LVs encoding shRNAs that specifically target only the aberrant HBB mRNA (shIVSI-110 Mid and Mid2) significantly increased the synthesis of HBB chains (HBB:Hba ratio) in transgenic MEL $HBB^{IVSI-110}$ clones, even though no knock-down of the aberrant mRNA was detected. These data suggest a possible alternative mechanism of silencing of the aberrant HBB mRNA by shIVSI-110 Mid/Mid2, such as via translational repression and/or mRNA destabilisation.

- Knock-down of the aberrant HBB mRNA was observed only in pLKO.I shIVSI-110 Down transduced MEL $HBB^{IVSI-110}$ clonal cells, which however led to a marginal increase of the HBB chain synthesis (HBB:Hba ratio).

- The excessive cell death and close to background average VCN in pLKO.I shIVSI-110 Up transduced MEL MA821 $HBB^{IVSI-110}$ clonal cells suggested that the expression of the particular shRNA could be deleterious to the murine cells.

- Transduction of patient-derived $CD34^+$ cells with pLKO.I shIVSI-110 Mid alone or in combination with MA821 LV (aka GLOBE) led to an increase of the HBB chains by 30-44% compared to (endogenous and/or vector-derived) background levels.

- Importantly, expression of shIVSI-110 Down and especially shIVSI-110 Mid and Mid2 in patient-derived $CD34^+$ increased the translation of the endogenous correctly spliced HBB mRNA to levels equal to or above those observed in MA821-transduced patient-derived $CD34^+$ cells with average VCN ~2.7.

- Transduction of patient-derived $CD34^+$ cells with pLKO.I shIVSI-110 Mid or Mid2 LVs led to a significant increase of the HBB chain synthesis (HBB:HBA) and of the percentage of late-stage erythroid differentiated cells (Ortho and Ret), relative to the untransduced negative controls. Both constitute essential indices for the correction of the β-thalassaemic phenotype of cells.

- In combination with MA821, application of shIVSI-110 Mid gives an additional, significant increase of HBB:HBA ratios compared to application of MA821 alone, with implications for the benefit of combination therapy for corresponding patients.

- In general, proof-of-principle of our hypothesis was established in both, humanized transgenic $HBB^{IVSI-110}$ clonal cells and IVSI-110 patient-derived $CD34^+$ cells, hence shRNA which specifically target the destruction of HBB aberrant mRNA can be used to improve the therapeutic potential of HBB LVs.

4.3 Personalized gene therapy by genome editing

As an alternative approach to gene augmentation via LVs, development of a personalised GT of the $HBB^{IVSI-110}$ misspliced mutation via genome editing was attempted. In this phase of the project a novel approach was established for the permanent functional correction of misspliced mutations, such as the $HBB^{IVSI-110}$ mutation, in which precise correction of the mutation is not required and restoration of splicing is achieved through the very efficient, common and error-prone NHEJ DNA repair pathway. Functional correction of misspliced mutations via the efficient NHEJ instead rather than inefficient HDR repair pathway, represents a promising application for the treatment of patient-derived $CD34^+$ cells, which are the main target of all the in-progress LV-based GT clinical trials for β-thalassaemia. On an aside, it has already been demonstrated that different primary mutations may be targeted by the same designer nuclease by HDR, given a donor template that would cover all mutations in question [214]. The observation indicates that not every mutation, including the sickling β-globin mutation, would require development of its own designer nuclease when employing NMD. However, detailed study of the efficiency of this approach is still required, and any gene-correction approach with appreciable efficiencies in primary HSPCs has so far relied on NHEJ.

The hypothesis behind the genome editing approach taken involves the restoration of correct splicing by disrupting the aberrant acceptor site and/or the adjacent upstream sequences, using specific designer nucleases (TALENs and RGEN). Screening the $HBB^{IVSI-110}$ sequence identified a unique target sequence for the RGEN, designed to specifically induced DSB at the site of the $HBB^{IVSI-110\ (G>A)}$ mutation and two target sites for TALEN pairs, in which they have a common binding site for the right TALEN monomer. The binding site of the HBB TALEN right monomer is immediately downstream of the $HBB^{IVSI-110(G>A)}$ mutation. Therefore the DSBs produced by the HBB TAL L1/R1 or L2/R1 pairs occurs upstream of the mutation, approximately in the middle of the two spacers, 10 bp and 13 bp, respectively. Even though the target sequence is within intron 1, which as suggested by others, *HBB* introns are better targets for genome editing than exons since they have less conserved homology sequences relative to *HBD*, a further increase in the specificity of the TALENs was attempted by gradually replacing the widely used the NN module that recognises G and A at lower affinity with NK, which specifically recognises only G but with lower affinity relative to NN [184,249]. Therefore, another three modified versions of the R1 monomer (R2, R3 and R4) in which 2, 4 and 6 NN modules were replaced by NK were produced.

In silico prediction of $HBB^{IVSI-110}$-specific RGEN off-targeting potential, as measured by MIT CRISPR design web tool, did not identify any gene with a high off-targeting risk. On the other

hand prediction of off-targeting for all the HBB TALEN pair combinations, using the PROGNOS web tool, designated *HBD* gene as a potential off-target (scores 82-76) whereas the on-target *HBB* score was 98. Noteworthy, scores against the *HBD* gene decreases with usage of the TALEN Right monomer with more NN>NK modifications.

Assessment of the disruption efficiency of the main designer nucleases, HBB TALEN pair R1/L1 and R1/L2, and the $HBB^{IVSI-110}$ RGEN *in vitro* on HEK293 cells, demonstrated the superiority of TALENs over RGEN for the on-target disruption of endogenous *HBB* by ~23% and ~10%, respectively. Assessment of the off-target potential in *HBD* on the same cell populations, showed close to background disruption levels with the highest off-targeting value (~4%) corresponding to HBB TALEN R1/L1. Although there is some off-targeting on the *HBD* gene, the disruption levels are still very low and thus unlikely to constitute a major drawback for the future applications in humans, since HbA_2 in adults comprises only 1.5-3.1% of the total haemoglobin in normal adults. Therefore, even in the case that *HBD* is disrupted in $hCD34^+$ cells, it would not be detrimental for the thalassaemia patients. According to the above data (4% *HBD* disruption), in the worst case scenario where all events that create a DSB at *HBD* lead to the disruption of the neighbouring *HBB* due to large deletion or inversions, this will reduce the therapeutic efficiency of the TALENs by only $1/5^{th}$.

Assessment of targeted disruption efficiency of all of our *HBB* targeting designer nucleases was performed in a more sensitive assay, using an episomal $HBB^{IVSI-110}$-GFP reporter construct, in which disruption efficiencies were measured by flow cytometry as a reduction in GFP fluorescence due to frame shifting of the eGFP cDNA ORF caused by DSB-induced INDELs. Again, the TALEN constructs proved to have a much higher target disruption efficiency than the RGEN, regardless the presence of the $HBB^{IVSI-110}$ mutation on the episomal GFP reporter, with the most efficient pairs (R1/L1 and L2/R1) exceeding the efficiency of the GFP specific nucleases, which were used as positive controls. In addition, the results obtained showed that the TALEN pairs combined with the more specific version of the right monomer (R2-4), as expected, had lower activity due to the lower binding affinity of the specific modules NK relative to NN for (G) guanine [182]. In particular, replacement of 2 out of 6 NN to NK modules reduced significantly the activity of both pairs by ~2/3, whereas further substitutions of 4 and 6 NN>NK modules, the reduction was not so profound. These data suggest that modifications of the N-terminus of the Right TALEN are more essential for the overall activity of the TALEN pair than those closer to the C-terminus. This is in agreement with previous studies which showed that the N-terminal TALE repeats contribute more to their affinity than the C terminal repeats [259]. However, these data are not conclusive as it is not known how the TALEN pairs would have behaved if the same NN>NK substitutions were made starting from the C terminus

to the N terminal end of the HBB Right TALEN monomer. Moreover, even though both sets with the unmodified pairs exhibited similar disruption efficiencies, all the HBB TALEN pairs with the modified right monomers (R2-4) combined with the L2 monomer indicated higher targeted disruption efficiency relative to the L1 counterpart pairs. This might be due to the better binding efficiency of the HBB TAL L2 to their binding sequence and/or the increased activity of the *Fok*I dimer due to the shorter spacer (10bp relative to 13bp).

4.3.1 *MEL MA821 HBB$^{IVSI-110}$ transgenic cell studies*

A proof of principle was established using the MEL MA821 HBB$^{IVSI-110}$ pool cells employing the most promising TALEN pairs (L1/R1 and L2/R1) and the RGEN. A full characterisation the type of INDELs produced was undertaken after genome editing with the two HBB TALEN pairs, R1/L1 and R1/L2, and RGEN by three different methods, T7E1 assay, TIDE and Sanger sequencing of TOPO clones. In general, TALEN pairs were more efficient than the RGEN and addition of TREX2 did not change the level of disruption or type of INDELs to a detectable degree, as others have shown in genome editing studies with meganucleases [250,260]. A likely explanation as to why TREX2 does not have the same affect in cells edited with TALENs and RGENs, is due to the different types of overhangs produced by each nuclease, mainly 5' overhangs and blunt ends, respectively [181,190,252]. In contrast, meganucleases produce 3' overhangs upon DSB, which are efficiently processed by the 3' to 5' prime exonuclease activity of TREX2, which mechanistically reduces the frequencies of DSBs, which re-ligate without losing any genetic material [190,250].

Characterisation of the INDELs by Sanger sequencing of TOPO clones of PCR products derived from genome edited MEL MA821 HBB$^{IVSI-110}$ VCN 1.9 cell populations indicated, as shown by others, that TALEN-induced INDELs mainly comprise deletions, which also is in agreement with the data obtained from TIDE analysis [181]. Furthermore, the TALEN pair with the larger spacer, 13 bp (R1/L1) tended to produce deletions starting from 3-4 bp, whereas the TALEN pair with the shorter spacer, 10 bp (R1/L2) produced deletions starting from 1-2 bp. One possible explanation for this observation is that the larger spacer between the two TALEN monomers, cause *Fok*I to create nicks on the gDNA, which are further apart, presumably as a result producing slightly larger overhangs. In general larger overhangs are more permissive for loss of more nucleotides than shorter overhangs before the re-joining the two DNA ends. A similar suggestion was made by Kim and Kim in an attempt to explain differences in mutation patterns in samples edited with TALENs and ZFNs, even though both use the same *Fok*I cleavage domain [181]. These authors showed that TALENs mainly produced deletion (89 % deletions versus 1.6 %) in mammalian cells, whereas ZFN-induced INDELs are more evenly distributed [181]. They

believed that the observed different mutation patterns between TALENs and ZFNs are due to differences in the length of spacers, in particular 12-21 bp and 5-6 bp, respectively. The larger spacer of TALENs possibly gives rise to a more heterogeneous overhangs than the defined 4 or 5 bp overhangs produced with ZFNs [181].

Most importantly, with the genome editing experiments on the transgenic MEL MA821 $HBB^{IVSI-110}$ VCN 1.9 cell pools, a proof-of-principle was established for the functional gene correction starting hypothesis at the mRNA and protein level. The results obtained are the first to demonstrate that targeted disruption of the aberrant splice acceptor site and/or the adjacent 5' polypyrimidine track region via genome editing, is able to restore the correct splicing pattern at levels which if translated to patient-derived $CD34^+$ cells would be therapeutic benefit. HBB-specific TALEN pairs, R1/L1 and R1/L2 were proven superior to the $HBB^{IVSI-110}$ RGEN system. In addition, due to a previous reported association of TREX2 with an increased risk of generating off-targeting [252], in combination with the observed relatively low contribution to gene disruption outcome, it was decided to excluded TREX2 from further experiments.

Data obtained from the correlation of specific type of INDELs with the level of restoration of pre-mRNA splicing and consequent increase in HBB chain production in different edited MEL MA821 $HBB^{IVSI-110}$ VCN 1 clones reinforced the functional gene correction starting hypothesis. Most importantly, the data obtained showed that disruption of the 5' region adjacent to the IVSI-110 mutation is sufficient to completely restore splicing, in which the mutation remains intact. This is to be expected as upstream from each acceptor splice site there are splicing elements (poly-pyrimidine tract, lariat branch point), which are essential for the recruitment and assembly of key spliceosome components especially the U2AF complex over the poly-pyrimidine tract and to promote general splicing efficiency [253]. The data obtained suggest that the size and position of the produced INDEL is crucial for the level of restoration of splicing and overall gene correction. Although there are a number elements and factors responsible for the regulation of splicing, the results of the experiments presented here indicate that the main elements that could be used as general targets for the functional correction of missplicing mutations through NHEJ, is primarily the mutation/aberrant splice site and the poly-pyrimidine tract. However, as there is no method available to control of the type and size of induced INDELs, careful design of the programmable nucleases should be performed, based on the mutation of interest. In the case of the $HBB^{IVSI-110}$ aberrant splice acceptor site was in close proximity with the normal acceptor site. On the one hand TALENs were designed to create a DSB 5' from the mutation, which reduced the risks of disrupting the normal acceptor splice site, poly-pyrimidine tract and coding sequence in exon II. On the other hand the $HBB^{IVSI-110}$

RGEN was designed to ensure the targeted disruption of the mutation but simultaneously increased the risk of disrupting the normal elements. The approach developed here provides a relatively wide window as a potential target sequence, allowing the selection of the most efficient designer nucleases with the minimum site effects (e.g., off-targeting or disruption of normal splicing).

The data obtained suggest that alteration of the poly-pyrimidine tract, especially removal of T and C, reduces the efficiency of pre-splicing complex formation on the aberrant splice acceptor site, hence forcing the spliceosome to use the normal acceptor site [81]. Another factor that possibly contributes to the preference of the spliceosome to use the normal acceptor site instead of the aberrant site in edited clones, is the decrease of the distance between the normal branch site and the 3‘ acceptor sites. In a previous study, it was shown that the abnormal splice acceptor site utilises the normal branch site, and therefore competes with the normal acceptor site [92]. In addition, Dominski and Kole were able to restore correct splicing by inactivating the normal branch site with specific antisense oligonucleotides [91]. This led to the activation of a cryptic branch site at nucleotide 107 of intron 1, which is in close proximity with the aberrant acceptor site, that was mainly utilized by the normal acceptor site reversing the ratios of aberrant/normal spliced products from 9:1 to 1:5 in treated samples [91]. Hence, most likely a minimum distance is also required between the branch site and 3’ acceptor splice site for the formation of a functional pre-spliceosome and the successful splicing of intron 1. The distance of the branch site from the 3’ aberrant and normal acceptor splice sites is 14 nt and 35 nt, respectively. Therefore, based on this hypothesis we believed that the complete restoration of splicing in edited clones where the aberrant splice acceptor site was left intact, is possibly due to the combination of disruption of the poly-pyrimidine tract and the decrease of the distance between the normal branch site and aberrant acceptor site by 4-9 bp, which in turn favours the use of the normal acceptor site. However, disruption of the aberrant acceptor is preferable since in all cases we observed restoration of full or close to normal pre-mRNA splicing and concomitant increase in translation and HBB chain production. The only exception was in clones where large deletions destroyed the normal acceptor splice site and exon 2 coding region.

4.3.2 *Patient$^{IVSI\text{-}110}$-derived CD34^{+} cell studies*

In general both TALEN pairs R1/L1 and R1/L2 performed equally in terms of targeted disruption efficiencies and restoration of splicing. The only minor difference was in the size of the induced deletion, where the R1/L1 pair showed a tendency to create deletions starting from 3-4 bp. Based on the correlation of INDELs with functional data, restoration of splicing

mainly occurs with deletions starting from 5bp, which also explains the observation that the majority of corrected clones are edited by the HBB TALEN R1/L1 pair. Based on these observations we used HBB TALEN R1/L1 in experiments on patient-derived $CD34^+$ cells. Although transfection by electroporation was optimised for the delivery of plasmid DNA in $hCD34^+$ cells in order to achieve high efficiencies (40-50%), the level of targeted disruption remained low (>12%). Unfortunately, this was not sufficient to allow the detection of restoration of splicing at the mRNA level, even though a marginal increase in the HBB:HBA chain ratio was detected. The results suggest that data obtained at the protein level are more reliable as a measure of genome editing events than those at the mRNA level. This in all likelihood is mainly because of the higher stability of the proteins relative to mRNA on the day of sample collection.

In addition to the general low targeted disruption efficiency, the high background HBB chain levels detected in the $hCD34^+$ liquid culture system, makes the assessment of the impact of splice restoration at the level of HBB chain synthesis more difficult and the HPLC method less sensitive to detect minor changes. Estimation of the HBB : HBA chain ratio in the edited cell sub-population relative to the non-edited, showed double the amount of HBB chains, which based on the already high background of HBB : HBA ratio (0.4), the edited cells have theoretically reached normal levels of expression (0.8 HBB : HBA chain ratio).

Thus based on these observations the proposed gene correction hypothesis is suitable for $patient^{IVSI\text{-}110}$-derived $CD34^+$ cells and the observed low targeted disruption efficiencies is not due to the low activity of HBB TALEN R1/L1 pair, but mainly due to technical and equipment limitations. Expression of the designer nucleases from plasmid DNA constructs is generally avoided, especially for editing cells as sensitive as $hCD34^+$ cells. The reduced efficiency of the TALEN pairs is most likely due to plasmid DNA-induced high cytotoxicity as well as low TALEN expression levels since plasmids require first to enter the nucleus to be expressed, which also delays the editing process. In addition, the high stability of plasmid DNA leads to a prolong expression of the nucleases, which increases the risk for off-target effects and genotoxicity.

Moreover, a conventional electroporator was employed, which is not designed for delivering designer nucleases to $hCD34^+$ cells. There are now commercially available new generation of electroporators, such as Nucleofectors (Lonza) or the Neon-electroporator system (Invitrogen), specifically designed with pre-set optimised protocols/conditions for genome editing of different types of cells, including $hCD34^+$ cells.

Nevertheless, despite the low efficiency of the genome editing tools employed here on patient-derived $CD34^+$ cells with the current equipment, a successful proof-of-principle was

established for the tested approach for the development of personalised GT of the *HBB*$^{IVSI-110}$ mutation. Promising preliminary data mainly on the transgenic cell pools and lines, demonstrates that high efficiency of complete restoration of splicing can be achieved via genome editing. One of the main advantages of the NHEJ-dependent GT approach presented over others seeking gene correction through HDR, is the fact that the target cells are hCD34^{+}, which are currently being used in clinically for allogeneic bone marrow transplantation and a number of GT clinical trials. Until recently most of the efforts aiming for gene correction of β-thalassaemia mutations were performed with iPSCs, which allows the selection and expansion of the corrected clones and their further differentiation to erythroid progenitors and finally to RBCs [211,214,216]. However, further studies should be performed on iPSCs for their long term characterisation before they are ready to be used in the clinic. The main barrier of gene correction in hCD34^{+} cells is the general low efficiency of HDR, in particular in HPSC cells, probably because of their general state of quiescence and slow cycling. However, recently a protocol has been described, which allows correction of the defective *IL2RG* gene in SCID-X1 patient bone marrow CD34^{+} cells via homology-directed gene targeting using ZFNs and donor template delivered as mRNA and an IDLV vector, respectively [218]. They achieved 3-11% homology-directed gene targeting of the treated cells depending on their primitive versus committed progenitor status [218]. The authors showed that more primitive cells are less proficient at performing HDR than committed progenitor cells since they divide at lower rates, and that still NHEJ repair is the preferential DSB repair mechanism [218]. More recently, two studies have demonstrated the correction of the SCD mutation in patient-derived bone marrow CD34^{+} cells at rates ~20% leading to the production of functional HbA at rates of 7% [217,219]. In the first study HDR-gene correction was achieved using ZFN mRNA and co-delivery of the DNA donor template via IDLV, whereas in the second, CRISPR and donor DNA template were delivered in the same cassette via IDLV and the Cas9 as mRNA by nucleofection [217,219]. Although it is still not clear as to why this is the case, in the second study the percentages of HDR-corrected cells (~20%) were higher than cells with INDELs (~16.3%) *in vitro* cultures. However, following transplantation in immunocompromised mice the rates of HDR-corrected cells dropped, whereas the NHEJ-edited cells were better maintained. These data suggest a higher efficiency of NHEJ repair in editing long-term repopulating HPSC than that achieved by HDR [219]. In general based on HSCT studies, 10-30% of bone marrow chimerism is sufficient to provide a clinical benefit in β-thalassaemia patients [209,261]. These exciting advances are bringing gene correction of HPSC via HDR closer to their clinical application, which can provide the correction of mutation defects, either by integrating the therapeutic expression cassette in safe-harbour sites (e.g., *AAVS1* or *CCR5* regions), or by insertion of a functional cDNA fragment downstream of an endogenous promoter. Complete correction of mutations is superior than

the former strategy as it offers full restoration of mutant gene function at physiological levels [218].

However, the genome editing approach developed here still has several advantages over the above HDR-based approaches in terms of therapeutic potential efficiency and safety, but with some drawbacks. Initially the proposed strategy is expected to be more efficient since it exploits the more efficient NHEJ DNA repair pathway, which in contrast to HDR, is generally active in all phases of the cell cycle, and most importantly in long term repopulating HPSC, which are also the main target of GT [262]. In addition, the use of the proposed NHEJ pathway does not require a donor DNA template. Therefore the presence of the designer nucleases in the same cells is sufficient for the functional correction of missplicing defects, increasing further its efficacy. Moreover, the presence of donor DNA in cells increases the risk of insertional mutagenesis, as part or even the whole DNA donor construct can be either randomly integrated in the genome, or more precisely inserted at nuclease off-target sites. In the absence of donor DNA, DSB would most likely be re-ligated without losing any genetic information. The main drawback of the genome editing approach proposed and developed here is that its mutation-specific and therefore different designer nucleases would need to be designed based on each case. However, since functional correction can be achieved with a range of INDELs adjacent to the mutation, this allows high flexibility in design and selection of the appropriate type of programmable nucleases suitable for each mutation. Finally, we suggest that the proposed personalised GT strategy for *HBB*$^{IVSI-110}$ β-thalassaemia can become a model, which can be adapted for the functional correction of human diseases caused in similar misspliced mutations.

4.3.3 *Summary of Conclusions: Genome editing approach*

- For the specific HBB$^{IVSI-110}$ target, TALEN pairs have proven superior genome-editing tools compared to RGEN in HEK293T cells, with low off-targeting (1/5th of the percentage of *HBB*-specific disruption) on the paralogous and highly sequence-similar *HBD* .

- TALEN pairs combined with the more specific version of the right monomer (R2-4), had lower activity due to the lower binding affinity of the specific modules NK relative to NN for (G) guanine

    - Replacement of 2 out of 6 NN to NK modules reduced significantly the activity of both pairs by ~2/3, whereas further substitutions of 4 and 6 NN to NK modules had a less profound impact. These data suggest that modifications of the N-terminus of

the Right TALEN are more essential for the overall activity of the TALEN pair than those closer to the C-terminus.

- Full characterisation of the type of INDELs produced by TALEN pairs (L1/R1 and L2/R1) and RGEN in humanised MEL MA821 $HBB^{IVSI-110}$ cell lines showed:
    - $HBB^{IVSI-110}$ TALEN pairs are superior genome editing tools compared to RGEN for the given target.
    - Addition of the TREX2 did not increase the level of targeted disruption efficiency of the designer nucleases, likely because TALENs and RGEN mainly produced 5' overhangs and blunt ends, respectively, which are unsuitable substrates for TREX2.
    - TALEN-induced INDELs mainly comprise deletions.
    - The TALEN pair with the larger, 13-bp spacer (R1/L1) tended to produce deletions starting from 3-4 bp, whereas the TALEN pair with the shorter, 10-bp spacer (R1/L2) produced deletions starting from 1-2 bp.
- Disruption of the region upstream of the aberrant acceptor site is also sufficient to restore correct splicing; however inclusion of the aberrant acceptor splice site in the disrupted region is generally more effective.
- In patient-derived $CD34^+$ cells, restoration of splicing at protein level was not detectable, owing to the low percentage of genome-edited patient-derived $CD34^+$ cells.
- In patient-derived $CD34^+$ cells, the low targeted disruption efficiency of the TALEN R1/L1 pair is probably the outcome of technical limitations (usage of conventional electroporator / delivered designer nucleases as plasmids) and not due to the TALEN design.
- In general, proof of principle was established, mainly in the transgenic cell pools and clonal lines, demonstrating that high efficiency of complete restoration of splicing can be achieved via genome editing.
- The proposed personalised GT strategy for $HBB^{IVSI-110}$ β-thalassaemia can become a model for the functional correction of other human diseases that are likewise caused by mutations that produce aberrant regulatory regions outside open reading frames.

5 Future plans

5.1 Gene addition

The final aim of the gene addition study is the optimisation of the therapeutic potential of *HBB*-based LV vectors, such as the MA821 HBB^{Normal} ("GLOBE"), in patients with thalassaemia misspliced mutations, which lead to the production of persistent non-functional HBB aberrant mRNA, such as $HBB^{IVSI-110}$.

Even though, the results obtained have shown an increase in translation of the correctly spliced HBB mRNA *in vitro* in MEL transgenic cells and patient-derived $CD34^+$ cells when transduced with pLKO.I shIVSI-110 Mid LVs, this was achieved at high VCNs. In addition, analysis of the knock-down effect at the RNA level in the transgenic cells did not indicate silencing of the aberrant HBB mRNA, regardless of the increase of translation in cells transduced with pLKO.I shIVSI-110 Mid/Mid2 LVs.

In the short-term, it is desirable to perform more experiments on patient-derived $CD34^+$ cells at lower VCNs. This will provide a better understanding of the actual role of the shIVSI-110 Mid RNA construct, as side-effects caused by the increased VCNs (high cell death, genome instability, de-regulation of RNAi pathways from shRNA over-expression) would be minimised [263]. The reduced toxicity will probably allow transduced $CD34^+$ cells cultures to reach numbers that will allow the collection of cells for RNA extraction at earlier stages of differentiation, providing more a reliable evaluation of aberrant mRNA silencing. In addition, for a more informative quantification of the two HBB variants, the combined method of absolute and relative quantification can be used, in which absolute quantities of each variant are normalised against the expression levels of the differentiation-normalised total HBB mRNA relative to the negative controls. This will provide extra information regarding the specificities and the effect of shIVSI-110 constructs on the aberrant mRNA.

In addition, based on the current data, the exact mechanism that is causing the increase in HBB chains in cells expressing the shIVSI-110 Mid is not clear. Support of the proposed hypothesis, could be achieved by comparison of polysomal profiles of MEL MA821 $HBB^{IVSI-110}$ cells and $patient^{IVSI-110}$-derived $CD34^+$ cells, transduced with pLKO.I shIVSI-110 Mid LVs relative to the negative controls (untransduced/shScramble). Analysis of polysome-bound RNAs will elucidate if the increase of HBB chains is due to increase of transcription of the endogenous gene (which could also be confirmed with the already proposed combined RT-qPCR method described herein) or due to an increase of the rate of translation, upon which the project hypothesis is based. In the case of the latter, it is expected that there will be a higher co-

localisation of the correctly spliced HBB mRNA with the polysomal fraction, which is associated with highest translational activity in samples expressing the shIVSI-110 Mid relative to negative controls.

A quick experiment, which will demonstrate more directly the silencing efficiency of the shIVSI-110 Mid/Mid2, is to perform an analogous experiment to that carried out on the episomal $HBB^{IVSI-110}$ GFP reporter construct for the assessment of targeted disruption efficiencies of the designer nucleases via flow cytometry. As mentioned in the Results chapter section 2.8.3, the $HBB^{IVSI-110}$ GFP reporter contains between the ATG start codon and eGFP cDNA the $HBB^{IVSI-110}$ intron 1 target sequence of the designer nucleases, which apparently includes and the part of the intron which is aberrantly spliced into the mature mRNA. In this case, the $HBB^{IVSI-110}$ eGFP reporter transcript represents an ideal target for the shIVSI-110 Mid/Mid2 constructs, as they both theoretically target the 5' end region of the $HBB^{IVSI-110}$ eGFP mRNA. Therefore, even though the target sequence is not in the context of the HBB mRNA sequence, co-transfection of the $HBB^{IVSI-110}$ GFP reporter and pLKO.U6 shIVSI-110 Mid/Mid2 constructs, can provide a quick evaluation of the knock-down efficiencies of shIVSI-110 Mid/Mid2 at the mRNA level (by RT-qPCR) and protein level (by flow cytometry). This will also allow an accurate and easy comparison of the efficiencies of the two shIVSI-110 Mid versions. In addition, this method can also constitute the platform for the fast and inexpensive assessment of the silencing efficiencies of the new optimised vectors, in which the shIVSI-110 Mid will be expressed in the context of miRNAs, either independently or as intronic miRNA co-expressed with MA821 HBB^{Normal} transgene.

The long-term plans are, as already mentioned above, is the expression of the shIVSI-110 Mid in the context of an miRNA structure (e.g., miR 30) under an RNA pol2 promoter, in which it has been shown to be less toxic and to not induce an interferon response [264]. The final goal is the development of an MA821 HBB miRIVSI-110 Mid LV, where the miR IVSI-110 Mid will be cloned in one of the two introns of *HBB* within the MA821 HBB^{Normal} transgene, at a site where it would not significantly affect the expression levels of the *HBB* transgene and at the same time allow the efficient processing of the intronic pre-miRNA to mature miRIVSI-110 Mid and the subsequent silencing of the aberrant HBB aberrant mRNA.

In addition, it is planned to test the original hypothesis of targeting aberrant mRNAs in patient derived $CD34^+$ cells harbouring similar intronic mutations (e.g., IVSII-745) to evaluate the reproducibility and consistency of the approach developed here with other mutations.

5.2 Genome editing

The main objective of this study was the development of a personalised GT approach for the *HBB*$^{IVSI-110}$ mutation via genome editing, using specific designer nucleases such TALENs and RGENs. The results obtained during this project clearly establish proof-of-principle in MEL MA821 HBB$^{IVSI-110}$ transgenic cell lines. However, application in patient-derived CD34^{+} cells was limited mainly due to the inefficient delivery and expression of the designer nucleases.

The main aim is the assessment of the therapeutic potential of the HBB TALEN R1/L1 genome editing tools developed here in patient derived CD34^{+} cells, using more efficient and less toxic delivery methods. The most attractive approach for delivering the HBB TALEN R1/L1 pair to patient-derived CD34^{+} cells is as mRNAs via nucleofection. This approach has many advantages over others, such as plasmid DNA and viral vectors, which makes it safer and more suitable to be used in a clinical context. Using mRNAs encoding the HBB TALEN R1/L1 pair avoids the use viral vectors, which still hold several limitations as the immunogenicity, limited DNA packaging capacity, especially for packaging the larger TALEN constructs, difficulties associated with the production of high-titre viral vectors, and the potential to lead to an undesired integration events into the host genome[265]. In addition, the limited packaging capacity of majority of the used viral vector fails to allow the delivery of both TALEN expression cassettes in one vector. Therefore, a system which requires the delivery of two viral vectors for genome editing may lead to increased toxic side effects and immune response, mainly due to the required high viral vector dose[266]. Furthermore, even though IDLVs are showed to have low immunological response, the delivery of TALEN constructs might lead to rearrangement of the TALE repeats due to recombination [266]. Most importantly, TALEN mRNAs provide a more efficient expression of the TALENs, as once they enter the cells they can immediately be translated to proteins. In contrast to viral and plasmid vectors, TALEN encoding mRNA results in dose-depended genome editing efficiencies, which in combination with their short half-life, genome editing can be better controlled in order to achieve monoallelic or biallelic disruption of the aberrant mutation reducing the risk of off-target effects. We have already transferred the TALEN constructs developed here into a plasmid construct, pPIX_K_TALEN_FokI, which allows their efficient *in vitro* transcription to mRNA. Preliminary experiments performed by collaborators with HBB TALEN R1/L1 pair mRNAs have achieved 95% targeted disruption of the *HBB* loci with ~10% off-targeting of *HBD* in *HBB*$^{IVSI-110}$ patient-derived CD34^{+} cells. Since a nucleofection instrument is not available in the host laboratory, direct future plans first includes the genome editing of patient-derived CD34^{+} cells with HBB TALEN R1/L1 pair mRNA and nucleofection, cell expansion and finally induced erythroid differentiation. This will allow the complete

assessment of the therapeutic potential of the approach and tools developed during this project, as well as their off-targeting potential.

Similarly, we would also like to increase the targeted disruption efficiency of our $HBB^{IVSI-110}$ RGEN on patient-derived $CD34^+$ by using more efficient delivery methods. The use of RGENs has the advantage that it is smaller than TALENs, and multiple sites can be genome edited with the delivery of multiple RGENs that utilise the same Cas9 (multiplexing) [267]. This can be used for simultaneous targeting of compound heterozygotes with misspliced mutations (e.g., IVSI-110/IVSII-745). RGEN tools should carefully be designed to avoid *in cis* off-targeting DSBs, which could lead to large deletions or inversions, whereas *in trans* on-targeting DSBs can lead to translocations, both resulting to the disruption of the mutant *HBB* genes instead of correcting them.

Before moving to a more complicated and risky strategy, we need to evaluate the correction levels on $CD34^+$ cells from compound heterozygote patients holding on one allele the IVSI-110 mutation. Genome editing of these cells will provide evidence showing if the functional correction of the one allele is sufficient to correct the thalassaemic phenotype (by reaching heterozygote expression levels).

The first alternative method to RGEN/Cas9 plasmid DNA electroporation is via nucleofection of Cas9 ribonucleoprotein complex ($HBB^{IVSI-110}$ gRNA RNA/Cas9 protein). The second alternative is the delivery via IDLV in which both RGEN and Cas9 are expressed from the same vector. Both methods are already widely used and constitute very efficient approaches for the delivery of RGENs to $hCD34^+$ and most importantly both are already established in our collaborator's laboratories.

Subsequent to the efficient delivery of our designer nucleases and assessment of functional correction on patient-derived $CD34^+$ cell pools *in vitro,* clonogenic assays of the edited cells could be performed in semi-solid methyl celluloses cultures for the evaluation of the multi-linage potential of edited cells.

One of the most essential parameters to measure is the evaluation of the off-targeting potential of the most promising tools from this project, namely HBB TAL R1/L1 and/or R1/L2, in $hCD34^+$ cells. Initially, tests can be conducted to evaluate disruption of the highly scored potential off-target sites identified by *in silico* web tools with the traditional gel-based detection methods, e.g., the T7E1 assay. However, this method is characterised with low detection sensitivity ~1-2% and is restricted to the sites that web tools identified as potential off-targets, hence the rest of the sites are overlooked [190]. For the more informative and deep analysis of off-targeting one can use sequencing-based methods, either Sanger sequencing or

next-generation sequencing, which provides higher sensitivity (down to 0.01% mutation frequencies) [190].

Further to the *in vitro* assessment of functional correction of the *HBB*$^{IVSI-110}$ mutation in patient-derived CD34^{+} cells, the next logical step is evaluation of the long-term repopulating potential of the corrected cells, with transplantation of genome edited patient-derived CD34^{+} cells into irradiate NOD SCID mice, and followed by secondary BM transplantations. This will provide a long-term assessment of the level of correction, engraftment and multi-potency potential of the edited CD34^{+} cells. Furthermore, the functional correction of the *HBB*$^{IVSI-110}$ mutation can also be assessed *in vivo* on the HBB$^{IVSI-110}$ humanised mouse model developed by Vadolas and colleagues, which recapitulates splicing defects seen in thalassaemic patients carrying the *HBB*$^{IVSI-110}$ mutation [90]. Application of the tools of this project using this mouse model system will provide the long-term characterisation of splice restoration at the mRNA and protein levels and evaluation of potential improvements in haematological abnormalities. Finally, the ability the genome editing tools of this project to correct via NHEJ in murine long-term HSCs should be undertaken.

Our final aim is the validation of the genome editing strategy developed here targeting other β-thalassaemia mutations, such as IVSII-745, as well as the full characterisation of the already developed tools (HBB TALENs) in terms of safety and correction efficiency that will presumably allow their use in clinical trials.

6 Appendix

6.1 In silico off-targeting analysis for TALENs - PROGNOS web tool

The hg19 genome was searched for the TALEN pair targeting:
5'- TTCTGATAGGCACTGACTC NN..NN GTCTATTTTCCCACCCTTA -3'
Using these RVDs:

HBB TALEN L1 (Left): 01NG02HD03NG04NN05NI06NG07NI08NN09NN10HD11NI12HD13NG14NN15NI16HD17NG18HD
HBB TALEN R1 (Right): 01NI02NI03NN04NN05NN06NG07NN08NN09NN10NI11NI12NI13NI14NG15NI16NN17NI18HD

Up to 6 mismatches were allowed in each nuclease half-site.

Allowed spacing distances were: 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Homodimers and heterodimers of the nucleases were included in the output

3260 total sites were located	
Mismatch Types	Genomic Regions
0 and 0: 1	Exons: 65
0 and 4: 1	Promoters: 21
3 and 5: 2	Introns: 1290
3 and 6: 8	Intergenic: 1884
4 and 5: 16	
4 and 6: 94	
5 and 5: 52	
5 and 6: 695	
6 and 6: 2391	

Table 12: In detail off-targeting analysis by PROGNOS web tool of HBB TALEN pair L1/R1.

Off target sites are ranked by the 'TALEN' algorithm. Rankings are given from low to high, so rank number 1 is the most likely predicted off-target site. Although ranking is based on TALEN score; "Homology" score is still given as a reference. In which higher homology scores indicate a more likely off-target site. Orientation of TALEN pairs on their target site sequence is shown with the predicted spacer length between the two monomers. Number of mismatches and their positions on the binding sequence of each monomer are designated with lower-cases font. Position of the predicted off target site on chromosome is also indicated. In addition the closed gene to the target site is listed as also and what genomic region is in (Exon/Intron/Promoter/Intergenic sequence).

PROGNOS analysis for TALEN pairs: HBB TALEN L1/R1												
Ranking	Homology Score	TALEN Score	Nuclease Orientation	Spacer Length	Mismatches	Left Half Site	Spacer Seq	Right Half Site	Chromosome	Chromosome Location	Genomic Region	Closest Gene
1	98	100	R-13-L	13	0_0	TAAGGGTGGGAAAATAGAC	caataggcagaga	GAGTCAGTGCCTATCAGAA	chr11	5248031	Intron	*HBB*
2	58	82.35	R-11-L	11	4_0	TgAGGGTaGGAAAAcAGcC	caagggacaga	GAGTCAGTGCCTATCAGAA	chr11	5255445	Intron	*HBD*
3	13	64.03	R-24-R	24	5_4	TAAGGaTGGagAAATAaAC	taatccctaatctctaacattatc	tTCTATTcTCaCACaCTcA	chr4	87773874	Intergenic	*SLC10A6*
4	10	62.73	R-28-R	28	6_4	aAAcGaTGGaAAAAgAaAC	aaattcagaacattgtcagcattctcat	GTCTtTTTTttCcCCCTTA	chr20	33925923	Intron	*UQCC*
5	10	62.59	R-11-R	11	6_4	TAAaaGTGtGAAcATtGAa	tgtatgtttca	GTCTtTTTTtCCAtCCcTA	chr8	88353583	Intron	*CNBD1*
6	13	62.53	R-19-R	19	5_4	TcAGGGaGGaAAAATAaAt	aagaggcatccattgtata	GTCTATTTTCaCAatgTTA	chr5	29741716	Intergenic	*LSP1P3*
7	10	62.47	R-26-R	26	6_4	TgAtGaTGGaAAAgcAGAC	tcaggcagaagtcttaacctttagga	GcCTATTTTCCCAtCtgTA	chr6	91222707	Intergenic	*MAP3K7*
8	20	62.07	L-14-R	14	5_3	aTCTGccAaGCACTGgCTC	ccaccttcagacca	GTCTATTTTCaCtgCCTTA	chr10	131159587	Intergenic	*MGMT*
9	10	61.68	R-13-L	13	6_4	TAAGGaaGGGAAAATAaAa	ttaaaaaaaaaca	ttGTtAGgGtgTATCAGAA	chr5	45353352	Intron	*HCN1*
10	20	61.51	R-30-R	30	5_3	TAAGGaTGGGAgAtaAGAt	agaagccaggtgtgcccgatggagctccct	GTCTtccTTCCCACCCTTA	chr10	24664139	Intron	*KIAA1217*
11	13	61.31	R-21-R	21	5_4	TAAaGGTGaGgAtATAGcC	tgatccctccccgttccagta	GcCaATTTTCCCAgCaTTA	chr8	91646379	Intron	*TMEM64*
12	5	61.18	R-28-R	28	6_5	TAAGaGgaGaAAAATAaAC	aaatacctaaggattagaaagaatataa	agCTATTTTCCaAtaCTgA	chr4	61821958	Intergenic	*LPHN3*
13	5	61.12	R-17-R	17	6_5	TAAaGaTaGtAAAAagGAC	ttgggaaatataaaata	cTCTATTTTCtCAgttTTA	chr8	107411475	Intron	*OXR1*
14	5	60.92	R-22-R	22	6_5	TAtGGaTGcaAAAATAaAa	aaaagcaaacacacctgtatta	aTtTATTTTCaCACttTTA	chr5	19782154	Intron	*CDH18*
15	10	60.9	L-27-R	27	6_4	TTCaGcTAGaCACTaACTC	acaaagtatgatacagtgtcttcattg	GgCccTcTTCtCAgCCTTA	chr22	27361583	Intergenic	*MIAT*

The hg19 genome was searched for the TALEN pair targeting:
5'- TTCTGATAGGCACTGACTC NN..NN GTCTATTTTCCCACCCTTA -3'

Using these RVDs:
HBB TALEN L1 (Left): 01NG02HD03NG04NN05NI06NG07NI08NN09NN10HD11NI12HD13NG14NN15NI16HD17NG18HD
HBB TALEN R2 (Right): 01NI02NI03NK04NK05NN06NG07NN08NN09NN10NI11NI12NI13NI14NG15NI16NN17NI18HD

Up to 6 mismatches were allowed in each nuclease half-site.

Allowed spacing distances were: 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Homodimers and heterodimers of the nucleases were included in the output

3260 total sites were located	
Mismatch Types	Genomic Regions
0 and 0: 1 0 and 4: 1 3 and 5: 2 3 and 6: 8 4 and 5: 16 4 and 6: 94 5 and 5: 52 5 and 6: 695 6 and 6: 2391	Exons: 65 Promoters: 21 Introns: 1290 Intergenic: 1884

Table 13: In detail off-targeting analysis by PROGNOS web tool of HBB TALEN pair L1/R2.

Off target sites are ranked by the 'TALEN' algorithm. Rankings are given from low to high, so rank number 1 is the most likely predicted off-target site. Although ranking is based on TALEN score; "Homology" score is still given as a reference. In which higher homology scores indicate a more likely off-target site. TALEN pair's orientation on target site sequence is shown with the predicted spacer length between the two monomers. Number of mismatched and their positions on the binding sequence of each monomer are designated with lower cases. Position of the predicted off target site on chromosome is also indicated. In addition the closed gene to the target site is listed as also and what genomic region is in (Exon/Intron/Promoter/Intergenic sequence).

PROGNOS analysis for TALEN pairs: HBB TALEN L1/R2												
Ranking	Homology Score	TALEN Score	Nuclease Orientation	Spacer Length	Mismatches	Left Half Site	Spacer Seq	Right Half Site	Chromosome Name	Chromosome Location	Genomic Region	Closest Gene
1	98	100	R-13-L	13	0_0	TAAGGGTGGGAAAATAGAC	caataggcagaga	GAGTCAGTGCCTATCAGAA	chr11	5248031	Intron	*HBB*
2	58	81.34	R-11-L	11	4_0	TgAGGGTaGGAAAAcAGcC	caagggacaga	GAGTCAGTGCCTATCAGAA	chr11	5255445	Intron	*HBD*
3	13	61.01	R-24-R	24	5_4	TAAGGaTGGagAAATAaAC	taatccctaatctctaacattatc	tTCTATTcTCaCACaCTcA	chr4	87773874	Intergenic	*SLC10A6*
4	20	60.71	L-14-R	14	5_3	aTCTGccAaGCACTGgCTC	ccaccttcagacca	GTCTATTTTCaCtgCCTTA	chr10	$1.31x10^8$	Intergenic	*MGMT*
5	10	60.53	R-13-L	13	6_4	TAAGGaaGGGAAAATAaAa	ttaaaaaaaaaca	ttGTtAGgGtgTATCAGAA	chr5	45353352	Intron	*HCN1*
6	10	59.65	L-27-R	27	6_4	TTCaGcTAGaCACTaACTC	acaaagtatgatacagtgtcttcattg	GgCccTcTTCtCAgCCTTA	chr22	27361583	Intergenic	*MIAT*
7	10	59.43	R-28-R	28	6_4	aAAcGaTGGaAAAAgAaAC	aaattcagaacattgtcagcattctcat	GTCTtTTTTttCcCCCTTA	chr20	33925923	Intron	*UQCC*
8	20	59.38	R-30-R	30	5_3	TAAGGaTGGGAgAtaAGAt	agaagccaggtgtgcccgatggagctccct	GTCTtccTTCCCACCCCTTA	chr10	24664139	Intron	*KIAA1217*
9	17	58.96	L-20-R	20	6_3	TTCTGgTgGGCACTGtCTC	actctatgaccatacctttt	GgCTAgTTTCtCctCCTTt	chr8	$1.4x10^8$	Intergenic	*KCNK9*
10	10	58.21	R-13-L	13	6_4	TggGGaTaGGAAAAcAGAa	cctgacagactga	GAGgCAGgGtCTATgAGAA	chr11	60788840	Intergenic	*CD6*
11	13	58.13	R-17-L	17	5_4	cAAGGGgaGGAAAtTAcAC	agaagtgtgggtcccag	GAGgCAGTGatTATCgGAA	chr18	2046630	Intergenic	*METTL4*
12	10	57.87	L-22-R	22	6_4	TTCaGgTAGGacCTagCTC	tatgttctctggtccctgggcg	GTagAgTTTCCCAtCCTTA	chr3	51645919	Intron	*RAD54L2*
13	13	57.82	L-27-R	27	5_4	TTtTGAgAaGCACTGAtcC	ctccctttcaaattttccttactattt	GTCTATTTcCCtAtCaTTA	chr1	84457888	Intron	*TTLL7*
14	17	57.67	L-30-R	30	6_3	TTCTGATAGGCACaGAaaC	ctcaagatatttggggatcttttggaagaca	GgCTATTTTgggACCtTTg	chr1	$2.25x10^8$	Intergenic	*DNAH14*
15	5	57.6	R-26-L	26	6_5	TgAGGaTGGaAAAATgaAC	attccaaaactctgatttaaagatct	tAtTtAGTGtCTgTgAGAA	chr6	20576328	Intron	*CDKAL1*

The hg19 genome was searched for the TALEN pair targeting:
5'- TTCTGATAGGCACTGACTC NN..NN GTCTATTTTCCCACCCTTA -3'

Using these RVDs:
HBB TALEN L1 (Left): 01NG02HD03NG04NN05NI06NG07NI08NN09NN10HD11NI12HD13NG14NN15NI16HD17NG18HD
HBB TALEN R3 (Right): 01NI02NI03NK04NK05NK06NG07NK08NN09NN10NI11NI12NI13NI14NG15NI16NN17NI18HD

Up to 6 mismatches were allowed in each nuclease half-site.

Allowed spacing distances were: 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Homodimers and heterodimers of the nucleases were included in the output

3260 total sites were located	
Mismatch Types	Genomic Regions
0 and 0: 1 0 and 4: 1 3 and 5: 2 3 and 6: 8 4 and 5: 16 4 and 6: 94 5 and 5: 52 5 and 6: 695 6 and 6: 2391	Exons: 65 Promoters: 21 Introns: 1290 Intergenic: 1884

Table 14: In detail off-targeting analysis by PROGNOS web tool of HBB TALEN pair L1/R3.

Off target sites are ranked by the 'TALEN' algorithm. Rankings are given from low to high, so rank number 1 is the most likely predicted off-target site. Although ranking is based on TALEN score; "Homology" score is still given as a reference. In which higher homology scores indicate a more likely off-target site. TALEN pair's orientation on target site sequence is shown with the predicted spacer length between the two monomers. Number of mismatched and their positions on the binding sequence of each monomer are designated with lower cases. Position of the predicted off target site on chromosome is also indicated. In addition the closed gene to the target site is listed as also and what genomic region is in (Exon/Intron/Promoter/Intergenic sequence).

PROGNOS analysis for TALEN pairs: HBB TALEN L1/R3												
Ranking	Homology Score	TALEN Score	Nuclease Orientation	Spacer Length	Mismatches	Left Half Site	Spacer Seq	Right Half Site	Chromosome Name	Chromosome Location	Genomic Region	Closest Gene
1	98	100	R-13-L	13	0_0	TAAGGGTGGGAAAATAGAC	caataggcagaga	GAGTCAGTGCCTATCAGAA	chr11	5248031	Intron	*HBB*
2	58	78.44	R-11-L	11	4_0	TgAGGGTaGGAAAAcAGcC	caagggacaga	GAGTCAGTGCCTATCAGAA	chr11	5255445	Intron	*HBD*
3	10	57.63	L-27-R	27	6_4	TTCaGcTAGaCACTaACTC	acaaagtatgatacagtgtcttcattg	GgCccTcTTCtCAgCCTTA	chr22	27361583	Intergenic	*MIAT*
4	20	57.62	L-14-R	14	5_3	aTCTGccAaGCACTGgCTC	ccaccttcagacca	GTCTATTTTCaCtgCCTTA	chr10	$1.31x10^8$	Intergenic	*MGMT*
5	10	56.89	R-13-L	13	6_4	TAAGGaaGGGAAAATAaAa	ttaaaaaaaaaca	ttGTtAGgGtgTATCAGAA	chr5	45353352	Intron	*HCN1*
6	17	56.25	L-20-R	20	6_3	TTCTGgTgGGCACTGtCTC	actctatgaccatacctttt	GgCTAgTTTCtCctCCTTt	chr8	$1.4x10^8$	Intergenic	*KCNK9*
7	17	56.13	L-22-L	22	6_3	TTCccAatGcCACTaACTC	tccaggccaaatcacaaataac	cAGTCtGTGCCTATgAGAA	chr2	$2.11x10^8$	Intron	*UNC80*
8	17	56.1	L-20-L	20	6_3	TTCTGAaAGGCcCTGACTg	gtgtatgcatatgtgtgtgt	GtGTgAtTatCTAaCAGAA	chr3	82943960	Intergenic	*GBE1*
9	13	55.94	R-24-R	24	5_4	TAAGGaTGGagAAATAaAC	taatccctaatctctaacattatc	tTCTATTcTCaCACaCTcA	chr4	87773874	Intergenic	*SLC10A6*
10	17	55.89	L-30-R	30	6_3	TTCTGATAGGCACaGAaaC	ctcaagatatttggggatcttttggaagaca	GgCTATTTTgggACCtTTg	chr1	$2.25x10^8$	Intergenic	*DNAH14*
11	20	55.8	R-30-R	30	5_3	TAAGGaTGGGAgAtaAGAt	agaagccaggtgtgcccgatggagctccct	GTCTtccTTCCCACCCTTA	chr10	24664139	Intron	*KIAA1217*
12	13	55.63	R-17-L	17	5_4	cAAGGGgaGGAAAtTAcAC	agaagtgtgggtcccag	GAGgCAGTGatTATCgGAA	chr18	2046630	Intergenic	*METTL4*
13	17	55.54	R-11-L	11	6_3	gAAaGcTGcaAAAATAGAt	atttttctggc	aAGTCAGTGCtTtTCAGAA	chr11	27738150	Intron	*BDNF*
14	10	55.36	L-20-R	20	6_4	TTCTtATgaGCACaaACTt	attttttcagtaactcaggt	GTCaATcTTtCCACtCTTA	chr4	$1.28x10^8$	Intergenic	*INTU*
15	13	55.32	L-10-L	10	5_4	gTCTGAagGcCACTGgCTC	tgtcagaagt	tgtTCAGTGCCTtTCAGAA	chr11	44529867	Intergenic	*CD82*

The hg19 genome was searched for the TALEN pair targeting:
5'- TTCTGATAGGCACTGACTC NN..NN GTCTATTTTCCCACCCTTA -3'

Using these RVDs:
HBB TALEN L1 (Left): 01NG02HD03NG04NN05NI06NG07NI08NN09NN10HD11NI12HD13NG14NN15NI16HD17NG18HD
HBB TALEN R4 (Right): 01NI02NI03NK04NK05NK06NG07NK08NK09NK10NI11NI12NI13NI14NG15NI16NN17NI18HD

Up to 6 mismatches were allowed in each nuclease half-site.

Allowed spacing distances were: 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Homodimers and heterodimers of the nucleases were included in the output

3260 total sites were located	
Mismatch Types	Genomic Regions
0 and 0: 1 0 and 4: 1 3 and 5: 2 3 and 6: 8 4 and 5: 16 4 and 6: 94 5 and 5: 52 5 and 6: 695 6 and 6: 2391	Exons: 65 Promoters: 21 Introns: 1290 Intergenic: 1884

Table 15: In detail off-targeting analysis by PROGNOS web tool of HBB TALEN pair L1/R4.

Off target sites are ranked by the 'TALEN' algorithm. Rankings are given from low to high, so rank number 1 is the most likely predicted off-target site. Although ranking is based on TALEN score; "Homology" score is still given as a reference. In which higher homology scores indicate a more likely off-target site. TALEN pair's orientation on target site sequence is shown with the predicted spacer length between the two monomers. Number of mismatched and their positions on the binding sequence of each monomer are designated with lower cases. Position of the predicted off target site on chromosome is also indicated. In addition the closed gene to the target site is listed as also and what genomic region is in (Exon/Intron/Promoter/Intergenic sequence).

PROGNOS analysis for TALEN pairs: HBB TALEN L1/R4												
Ranking	Homology Score	TALEN Score	Nuclease Orientation	Spacer Length	Mismatches	Left Half Site	Spacer Seq	Right Half Site	Chromosome Name	Chromosome Location	Genomic Region	Closest Gene
1	98	100	R-13-L	13	0_0	TAAGGGTGGGAAAATAGAC	caataggcagaga	GAGTCAGTGCCTATCAGAA	chr11	5248031	Intron	*HBB*
2	58	76.83	R-11-L	11	4_0	TgAGGGTaGGAAAAcAGcC	caagggacaga	GAGTCAGTGCCTATCAGAA	chr11	5255445	Intron	*HBD*
3	17	56.13	L-22-L	22	6_3	TTCccAatGcCACTaACTC	tccaggccaaatcacaaataac	cAGTCtGTGCCTATgAGAA	chr2	$2.11x10^8$	Intron	*UNC80*
4	17	56.1	L-20-L	20	6_3	TTCTGAaAGGCcCTGACTg	gtgtatgcatatgtgtgtgt	GtGTgAtTatCTAaCAGAA	chr3	82943960	Intergenic	*GBE1*
5	10	55.6	R-13-L	13	6_4	TAAGGaaGGGAAAATAaAa	ttaaaaaaaaaca	ttGTtAGgGtgTATCAGAA	chr5	45353352	Intron	*HCN1*
6	13	55.32	L-10-L	10	5_4	gTCTGAagGcCACTGgCTC	tgtcagaagt	tgtTCAGTGCCTtTCAGAA	chr11	44529867	Intergenic	*CD82*
7	10	55.19	L-27-R	27	6_4	TTCaGcTAGaCACTaACTC	acaaagtatgatacagtgtcttcattg	GgCccTcTTCtCAgCCTTA	chr22	27361583	Intergenic	*MIAT*
8	8	55.1	L-26-L	26	5_5	TTtTtATAaGCACaGtCTC	tcccctccattgatgccaatggttt	cAGcCAGTGCtTATtAaAA	chr1	90858848	Intergenic	*BARHL2*
9	20	54.92	L-14-R	14	5_3	aTCTGccAaGCACTGgCTC	ccaccttcagacca	GTCTATTTTCaCtgCCTTA	chr10	$1.31x10^8$	Intergenic	*MGMT*
10	10	54.73	L-19-L	19	6_4	cTCTGATAGaCAaTaAggC	attgtacacctataagata	GgGTtAcTtCCTATCAGAA	chr4	$1.8x10^8$	Intergenic	*LINC00290*
11	5	54.6	L-18-L	18	6_5	TgCTGgTAGcCACTGgCaC	acattgttggccctccct	GAGTttGTGgtTAaCtGAA	chr5	$1.52x10^8$	Intergenic	*GLRA1*
12	13	54.18	R-17-L	17	5_4	cAAGGGgaGGAAAtTAcAC	agaagtgtgggtcccag	GAGgCAGTGatTATCgGAA	chr18	2046630	Intergenic	*METTL4*
13	17	53.83	L-20-R	20	6_3	TTCTGgTgGGCACTGtCTC	actctatgaccatacctttt	GgCTAgTTTCtCctCCTTt	chr8	$1.4x10^8$	Intergenic	*KCNK9*
14	5	53.63	L-20-L	20	6_5	TTCTGcaAcGCACTGcCTa	gtgttgtgcctggcagaaag	GgGTttGTGCtTAaCAaAA	chr22	35476976	Intron	*ISX*
15	17	53.52	L-30-R	30	6_3	TTCTGATAGGCACaGAaaC	ctcaagatatttggggatctttggaagaca	GgCTATTTTgggACCtTTg	chr1	$2.25x10^8$	Intergenic	*DNAH14*

The hg19 genome was searched for the TALEN pair targeting:
5'- TGATAGGCACTGACTCTCT NN..NN GTCTATTTTCCCACCCTTA -3'

Using these RVDs:
HBB TALEN L2 (Left): 01NN02NI03NG04NI05NN06NN07HD08NI09HD10NG11NN12NI13HD14NG15HD16NG17HD18NG
HBB TALEN R1 (Right): 01NI02NI03NN04NN05NN06NG07NN08NN09NN10NI11NI12NI13NI14NG15NI16NN17NI18HD

Up to 6 mismatches were allowed in each nuclease half-site.

Allowed spacing distances were: 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Homodimers and heterodimers of the nucleases were included in the output

4303 total sites were located	
Mismatch Types	Genomic Regions
0 and 0: 1 0 and 4: 1 3 and 5: 3 4 and 4: 1 3 and 6: 7 4 and 5: 18 4 and 6: 128 5 and 5: 59 5 and 6: 1112 6 and 6: 2973	Exons: 72 Promoters: 35 Introns: 1643 Intergenic: 2553

Table 16: In detail off-targeting analysis by PROGNOS web tool of HBB TALEN pair L2/R1.

Off target sites are ranked by the 'TALEN' algorithm. Rankings are given from low to high, so rank number 1 is the most likely predicted off-target site. Although ranking is based on TALEN score; "Homology" score is still given as a reference. In which higher homology scores indicate a more likely off-target site. TALEN pair's orientation on target site sequence is shown with the predicted spacer length between the two monomers. Number of mismatched and their positions on the binding sequence of each monomer are designated with lower cases. Position of the predicted off target site on chromosome is also indicated. In addition the closed gene to the target site is listed as also and what genomic region is in (Exon/Intron/Promoter/Intergenic sequence).

PROGNOS analysis for TALEN pairs: HBB TALEN L2/R1												
Ranking	Homology Score	TALEN Score	Nuclease Orientation	Spacer Length	Mismatches	Left Half Site		Right Half Site	Chromosome Name	Chromosome Location	Genomic Region	Closest Gene
1	98	100	R-10-L	10	0_0	TAAGGGTGGGAAAATAGAC	caataggcag	AGAGAGTCAGTGCCTATCA	chr11	5248031	Intron	*HBB*
2	58	82.35	R-8-L	8	4_0	TgAGGGTaGGAAAAcAGcC	caagggac	AGAGAGTCAGTGCCTATCA	chr11	5255445	Intron	*HBD*
3	13	64.03	R-24-R	24	5_4	TAAGGaTGGagAAATAaAC	taatccctaatctctaacattatc	tTCTATTcTCaCACaCTcA	chr4	87773874	Intergenic	*SLC10A6*
4	10	62.73	R-28-R	28	6_4	aAAcGaTGGaAAAAgAaAC	aaattcagaacattgtcagcattctcat	GTCTtTTTTttCcCCCTTA	chr20	33925923	Intron	*UQCC*
5	10	62.59	R-11-R	11	6_4	TAAaaGTGtGAAcATtGAa	tgtatgtttca	GTCTtTTTTtCCAtCCcTA	chr8	88353583	Intron	*CNBD1*
6	13	62.53	R-19-R	19	5_4	TcAGGGaGGaAAAATAaAt	aagaggcatccattgtata	GTCTATTTTCaCAatgTTA	chr5	29741716	Intergenic	*LSP1P3*
7	10	62.47	R-26-R	26	6_4	TgAtGaTGGaAAAgcAGAC	tcaggcagaagtcttaacctttagga	GcCTATTTTCCCAtCtgTA	chr6	91222707	Intergenic	*MAP3K7*
8	20	61.76	L-21-R	21	5_3	aGATAGGCcCTGACCcCTCT	ctgggttaatgactgtgctca	GTCTATTTTCtCttCCTat	chr2	68144909	Intergenic	*C1D*
9	20	61.51	R-30-R	30	5_3	TAAGGaTGGGAgAtaAGAt	agaagccaggtgtgcccgatggagctccct	GTCTtccTTCCCACCCTTA	chr10	24664139	Intron	*KIAA1217*
10	13	61.31	R-21-R	21	5_4	TAAaGGTGaGgAtATAGcC	tgatccctccccgttccagta	GcCaATTTTCCCAgCaTTA	chr8	91646379	Intron	*TMEM64*
11	5	61.18	R-28-R	28	6_5	TAAGaGgaGaAAAATAaAC	aaatacctaaggattagaaagaatataa	agCTATTTTCCaAtaCTgA	chr4	61821958	Intergenic	*LPHN3*
12	5	61.12	R-17-R	17	6_5	TAAaGaTaGtAAAAagGAC	ttgggaaatataaaata	cTCTATTTTCtCAgttTTA	chr8	$1.07x10^{8}$	Intron	*OXR1*
13	18	61.1	L-10-R	10	4_4	gGATAcaCACTGACTtTCT	tagagtctag	GTtTATTTTCCaAaCaTTA	chr9	73772020	Intergenic	*TRPM3*
14	20	61.07	L-14-R	14	5_3	TacTgGGtACTGACTtTCT	cctgaatctcgatt	GTCTATgTTCCCAtCCaTA	chr3	19547576	Intron	*KCNH8*
15	17	61.05	L-21-R	21	6_3	TtATtGGaAaTGcCTCTgT	taattgatgggtctctgatt t	GTCTATTTTCCCcCttTTA	chr2	4053770	Intergenic	*LOC100505964*

The hg19 genome was searched for the TALEN pair targeting:

5'- TGATAGGCACTGACTCTCT NN..NN GTCTATTTTCCCACCCTTA -3'

Using these RVDs:
HBB TALEN L2 (Left): 01NN02NI03NG04NI05NN06NN07HD08NI09HD10NG11NN12NI13HD14NG15HD16NG17HD18NG
HBB TALEN R2 (Right): 01NI02NI03NK04NK05NN06NG07NN08NN09NN10NI11NI12NI13NI14NG15NI16NN17NI18HD

Up to 6 mismatches were allowed in each nuclease half-site.

Allowed spacing distances were: 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Homodimers and heterodimers of the nucleases were included in the output

4303 total sites were located	
Mismatch Types	Genomic Regions
0 and 0: 1 0 and 4: 1 3 and 5: 3 4 and 4: 1 3 and 6: 7 4 and 5: 18 4 and 6: 128 5 and 5: 59 5 and 6: 1112 6 and 6: 2973	Exons: 72 Promoters: 35 Introns: 1643 Intergenic: 2553

Table 17: In detail off-targeting analysis by PROGNOS web tool of HBB TALEN pair L2/R2.

Off target sites are ranked by the 'TALEN' algorithm. Rankings are given from low to high, so rank number 1 is the most likely predicted off-target site. Although ranking is based on TALEN score; "Homology" score is still given as a reference. In which higher homology scores indicate a more likely off-target site. TALEN pairs orientation on target site sequence is shown with the predicted spacer length between the two monomers. Number of mismatched and their positions on the binding sequence of each monomer are designated with lower cases. Position of the predicted off target site on chromosome is also indicated. In addition the closed gene to the target site is listed as also and what genomic region is in (Exon/Intron/Promoter/Intergenic sequence).

PROGNOS analysis for TALEN pairs: HBB TALEN L2/R2												
Ranking	Homology Score	TALEN Score	Nuclease Orientation	Spacer Length	Mismatches	Left Half Site		Right Half Site	Chromosome Name	Chromosome Location	Genomic Region	Closest Gene
1	98	100	R-10-L	10	0_0	TAAGGGTGGGAAAATAGAC	caataggcag	AGAGAGTCAGTGCCTATCA	chr11	5248031	Intron	*HBB*
2	58	81.34	R-8-L	8	4_0	TgAGGGTaGGAAAAcAGcC	caagggac	AGAGAGTCAGTGCCTATCA	chr11	5255445	Intron	*HBD*
3	13	61.01	R-24-R	24	5_4	TAAGGaTGGagAAATAaAC	taatccctaatctctaacattatc	tTCTATTcTCaCACaCTcA	chr4	87773874	Intergenic	*SLC10A6*
4	20	60.61	L-21-R	21	5_3	aGATAGGCcCTGACCcCTCT	ctgggttaatgactgtgctca	GTCTATTTTCtCttCCTat	chr2	68144909	Intergenic	*C1D*
5	17	59.64	L-11-R	11	6_3	TGccAaGCACTGgCTCcCa	ccttcagacca	GTCTATTTTCaCtgCCTTA	chr10	$1.31x10^{8}$	Intergenic	*MGMT*
6	20	59.44	L-14-R	14	5_3	TacTgGGtACTGACTtTCT	cctgaatctcgatt	GTCTATgTTCCCAtCCaTA	chr3	19547576	Intron	*KCNH8*
7	10	59.43	R-28-R	28	6_4	aAAcGaTGGaAAAAgAaAC	aaattcagaacattgtcagcattctcat	GTCTtTTTTttCcCCCTTA	chr20	33925923	Intron	*UQCC*
8	20	59.38	R-30-R	30	5_3	TAAGGaTGGGAgAtaAGAt	agaagccaggtgtgcccgatggagctccct	GTCTtccTTCCCACCCTTA	chr10	24664139	Intron	*KIAA1217*
9	13	59.31	R-14-L	14	5_4	TgAGGaTGGGAAAATAaAg	gaacttaacaatgg	AGAGAcTttGTGaCTATtA	chr3	$1.64x10^{8}$	Intergenic	*SI*
10	17	58.63	L-8-L	8	6_3	TaATAGaCACaGACTggCa	aattggat	AaAGAGTCAGgGCCcATCA	chrY	15836005	Intergenic	*TMSB4Y*
11	18	58.31	L-10-R	10	4_4	gGATAcaCACTGACTtTCT	tagagtctag	GTtTATTTTCCaAaCaTTA	chr9	73772020	Intergenic	*TRPM3*
12	10	57.6	R-11-R	11	6_4	TAAaaGTGtGAAcATtGAa	tgtatgtttca	GTCTtTTTTtCCAtCCcTA	chr8	88353583	Intron	*CNBD1*
13	10	57.53	L-20-R	20	6_4	TGATgaaCACacACTCaCT	atgtccaactattattacta	GTaTtTcTTaCCACCCTTA	chr11	34335339	Intron	*ABTB2*
14	13	57.52	R-19-R	19	5_4	TcAGGGaGGaAAAATAaAt	aagaggcatccattgtata	GTCTATTTTCaCAatgTTA	chr5	29741716	Intergenic	*LSP1P3*
15	17	57.46	L-12-R	12	6_3	TGAcAtGgAtTGACTgTCa	atggtaaagttt	GTCTATTTTCCagtCCTTA	chr12	$1.12x10^{8}$	Exon	*ERP29*

The hg19 genome was searched for the TALEN pair targeting:
5'- TGATAGGCACTGACTCTCT NN..NN GTCTATTTTCCCACCCTTA -3'

Using these RVDs:
HBB TALEN L2 (Left): 01NN02NI03NG04NI05NN06NN07HD08NI09HD10NG11NN12NI13HD14NG15HD16NG17HD18NG
HBB TALEN R3 (Right): 01NI02NI03NK04NK05NK06NG07NK08NN09NN10NI11NI12NI13NI14NG15NI16NN17NI18HD

Up to 6 mismatches were allowed in each nuclease half-site.

Allowed spacing distances were: 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Homodimers and heterodimers of the nucleases were included in the output

4303 total sites were located	
Mismatch Types	Genomic Regions
0 and 0: 1 0 and 4: 1 3 and 5: 3 4 and 4: 1 3 and 6: 7 4 and 5: 18 4 and 6: 128 5 and 5: 59 5 and 6: 1112 6 and 6: 2973	Exons: 72 Promoters: 35 Introns: 1643 Intergenic: 2553

Table 18: In detail off-targeting analysis by PROGNOS web tool of HBB TALEN pair L2/R3.

Off target sites are ranked by the 'TALEN' algorithm. Rankings are given from low to high, so rank number 1 is the most likely predicted off-target site. Although ranking is based on TALEN score; "Homology" score is still given as a reference. In which higher homology scores indicate a more likely off-target site. TALEN pair's orientation on target site sequence is shown with the predicted spacer length between the two monomers. Number of mismatched and their positions on the binding sequence of each monomer are designated with lower cases. Position of the predicted off target site on chromosome is also indicated. In addition the closed gene to the target site is listed as also and what genomic region is in (Exon/Intron/Promoter/Intergenic sequence).

PROGNOS analysis for TALEN pairs: HBB TALEN L2/R3												
Ranking	Homology Score	TALEN Score	Nuclease Orientation	Spacer Length	Mismatches	Left Half Site		Right Half Site	Chromosome Name	Chromosome Location	Genomic Region	Closest Gene
1	98	100	R-10-L	10	0_0	TAAGGGTGGGAAAATAGAC	caataggcag	AGAGAGTCAGTGCCTATCA	chr11	5248031	Intron	*HBB*
2	58	78.44	R-8-L	8	4_0	TgAGGGTaGGAAAAcAGcC	caagggac	AGAGAGTCAGTGCCTATCA	chr11	5255445	Intron	*HBD*
3	17	58.63	L-8-L	8	6_3	TaATAGaCACaGACTggCa	aattggat	AaAGAGTCAGgGCCcATCA	chrY	15836005	Intergenic	*TMSB4Y*
4	20	57.7	L-21-R	21	5_3	aGATAGGCcCTGACCcCTCT	ctgggttaatgactgtgctca	GTCTATTTTCtCttCCTat	chr2	68144909	Intergenic	*C1D*
5	5	56.82	L-9-L	9	6_5	TaATAGtCAaaaACTCTCa	ggccttcca	AGAGAGTtttTGaCTATtA	chr6	$1.05x10^8$	Intergenic	*HACE1*
6	17	56.56	L-11-R	11	6_3	TGccAaGCACTGgCTCcCa	ccttcagacca	GTCTATTTTCaCtgCCTTA	chr10	$1.31x10^8$	Intergenic	*MGMT*
7	10	56.38	L-20-R	20	6_4	TGATgaaCACacACTCaCT	atgtccaactattattacta	GTaTtTcTTaCCACCCTTA	chr11	34335339	Intron	*ABTB2*
8	13	56.28	R-14-L	14	5_4	TgAGGaTGGGAAAATAaAg	gaacttaacaatgg	AGAGAcTttGTGaCTATtA	chr3	$1.64x10^8$	Intergenic	*SI*
9	20	56.16	L-14-R	14	5_3	TacTgGGtACTGACTtTCT	cctgaatctcgatt	GTCTATgTTCCCAtCCaTA	chr3	19547576	Intron	*KCNH8*
10	13	55.94	R-24-R	24	5_4	TAAGGaTGGagAAATAaAC	taatccctaatctctaacattatc	tTCTATTcTCaCACaCTcA	chr4	87773874	Intergenic	*SLC10A6*
11	10	55.9	L-8-L	8	6_4	TaAaAGaCACaGACTCgCa	aattggat	AaAGAGTCAagGCCcATCA	chr3	89883036	Intergenic	*EPHA3*
12	20	55.8	R-30-R	30	5_3	TAAGGaTGGGAgAtaAGAt	agaagccaggtgtgcccgatggagctccct	GTCTtccTTCCCACCCTTA	chr10	24664139	Intron	*KIAA1217*
13	13	55.73	L-8-L	8	5_4	TGATgGGCctTGACTCTtT	attcaatt	tGAcAGTCtGTGCCTtTtA	chr5	19078719	Intergenic	*CDH18*
14	8	55.62	L-30-L	30	5_5	TagaAatCACTGACTCTCT	tctaccacatactattggtcaaataaaatc	AtAGAGTCAGTtCaatTCA	chr5	12459185	Intergenic	*TAG*
15	10	55.61	L-23-L	23	6_4	TGAgAaaCACTtACatTCT	gcaaactaggtacaagtgcagct	AagaAGTCAGTGCCTATtA	chr2	58413085	Intron	*FANCL*

The hg19 genome was searched for the TALEN pair targeting:
5'- TGATAGGCACTGACTCTCT NN..NN GTCTATTTTCCCACCCTTA -3'

Using these RVDs:
HBB TALEN L2 (Left): 01NN02NI03NG04NI05NN06NN07HD08NI09HD10NG11NN12NI13HD14NG15HD16NG17HD18NG
HBB TALEN R4 (Right): 01NI02NI03NK04NK05NK06NG07NK08NK09NK10NI11NI12NI13NI14NG15NI16NN17NI18HD

Up to 6 mismatches were allowed in each nuclease half-site.

Allowed spacing distances were: 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Homodimers and heterodimers of the nucleases were included in the output

4303 total sites were located	
Mismatch Types	Genomic Regions
0 and 0: 1 0 and 4: 1 3 and 5: 3 4 and 4: 1 3 and 6: 7 4 and 5: 18 4 and 6: 128 5 and 5: 59 5 and 6: 1112 6 and 6: 2973	Exons: 72 Promoters: 35 Introns: 1643 Intergenic: 2553

Table 19: In detail off-targeting analysis by PROGNOS web tool of HBB TALEN pair L2/R4.

Off target sites are ranked by the 'TALEN' algorithm. Rankings are given from low to high, so rank number 1 is the most likely predicted off-target site. Although ranking is based on TALEN score; "Homology" score is still given as a reference, in which higher homology scores indicate a more likely off-target site. TALEN-pair orientation on the target site sequence is shown with the predicted spacer length between both monomers. The number of mismatches and their positions on the binding sequence of each monomer are designated with lower-case font. Position of the predicted off target site on chromosome is also indicated. In addition the closed gene to the target site is listed as also and what genomic region is in (Exon/Intron/Promoter/Intergenic sequence).

PROGNOS analysis for TALEN pairs: HBB TALEN L2/R4												
Ranking	**Homology Score**	**TALEN Score**	**Nuclease Orientation**	**Spacer Length**	**Mismatches**	**Left Half Site**		**Right Half Site**	**Chromosome Name**	**Chromosome Location**	**Genomic Region**	**Closest Gene**
1	98	100	R-10-L	10	0_0	TAAGGGTGGGAAAATAGAC	caataggcag	AGAGAGTCAGTGCCTATCA	chr11	5248031	Intron	*HBB*
2	58	76.83	R-8-L	8	4_0	TgAGGGTaGGAAAAcAGcC	caagggac	AGAGAGTCAGTGCCTATCA	chr11	5255445	Intron	*HBD*
3	17	58.63	L-8-L	8	6_3	TaATAGaCACaGACTggCa	aattggat	AaAGAGTCAGgGCCcATCA	chrY	15836005	Intergenic	*TMSB4Y*
4	5	56.82	L-9-L	9	6_5	TaATAGtCAaaaACTCTCa	ggccttcca	AGAGAGTtttTGaCTATtA	chr6	$1.05x10^8$	Intergenic	*HACE1*
5	10	55.9	L-8-L	8	6_4	TaAaAGaCACaGACTCgCa	aattggat	AaAGAGTCAagGCCcATCA	chr3	89883036	Intergenic	*EPHA3*
6	13	55.73	L-8-L	8	5_4	TGATgGGCctTGACTCTtT	attcaatt	tGAcAGTCtGTGCCTtTtA	chr5	19078719	Intergenic	*CDH18*
7	8	55.62	L-30-L	30	5_5	TagaAatCACTGACTCTCT	tctaccacatactattggtcaaataaaatc	AtAGAGTCAGTtCaatTCA	chr5	12459185	Intergenic	*TAG*
8	10	55.61	L-23-L	23	6_4	TGAgAaaCACTtACatTCT	gcaaactaggtacaagtgcagct	AagaAGTCAGTGCCTATtA	chr2	58413085	Intron	*FANCL*
9	10	55.56	L-16-L	16	6_4	TaAaAGaCACaGACTCaCa	gactggcaaattcgtt	AaAGAGTCAagaCCTATCA	chr10	49581467	Intergenic	*MAPK8*
10	10	55.38	L-8-L	8	6_4	TGATgtGtctTGACTCTtT	atccaata	tGAGAGTCtGTGCCTtTtA	chrX	23445342	Intergenic	*PTCHD1*
11	17	55.27	L-8-L	8	6_3	TGATgGGCACTGggTCTCT	cttcctcc	AGgGAaTCAGgGCCctcCA	chr21	46917484	Intron	*COL18A1*
12	5	55.2	L-29-L	29	6_5	TGcTAGGgACaacCTCTCc	acctcagggacaggttggcattctctaat	AGgGAGTtAGTGgtTAgCA	chr17	56548387	Intron	*HSF5*
13	20	55.08	L-21-R	21	5_3	aGATAGGCcCTGACCcCTCT	ctgggttaatgactgtgctca	GTCTATTTTCtCttCCTat	chr2	68144909	Intergenic	*C1D*
14	13	54.99	R-14-L	14	5_4	TgAGGaTGGGAAAATAaAg	gaacttaacaatgg	AGAGAcTttGTGaCTATtA	chr3	$1.64x10^8$	Intergenic	*SI*
15	10	54.91	L-8-L	8	6_4	TGATgGGtAtTGACTCTtT	atccaatt	tGcGAGTCtGTGtCTgTtA	chrX	$1.38x10^8$	Intron	*FGF13*

6.2 In silico off-targeting for RGEN - CRISPR Design MIT web tool

Table 20: HBB$^{IVSI1-110}$ RGEN in silico off-targeting prediction.
General analysis of HBB$^{IVSI-110}$ RGEN quality and off targeting potential on the human genome (hg19) based on CRISPR design MIT web tool.. Analysis includes RGEN quality score and the number of all potential off target site in the human genome as also and the site that lies within exons.

The hg19 genome was searched for the CRISPR targeting:

HBB$^{IVSI-110}$ RGEN	5' to 3' RGEN sequence- NGG (PAM)	quality score	on-target locus	number of off-target sites
	GGGTGGGAAAATAGACTAAT-NGG	55 (high quality)	*HBB$^{IVSI-110(G>A)}$*	264 (15 are in genes)

Table 21: In detail analysis of the off-targeting potential of HBB$^{IVSI-110}$ RGEN (exonic regions only).
Exonic off-target sites are ranked from low to high, so that rank number 1 is the most likely predicted off-target site based on the off-targeting score (highest to low). The *Ranking* column shows the position of the exonic off-target site in relation to all predicted off-target sites (265). Information of the exonic off-target sites is shown (e.g. exonic off-target sequence, name and UCSC gene and locus of the off-target site.). Numbers of mismatched bases (MM) between the HBB$^{IVSI-110}$ RGEN sequence and off-target sites are indicated as also their position on the RGEN sequence (5' to 3')

CRISPR design off-targeting analysis

Ranking	Exonic off-target sequence	Score	Mismatches	UCSC gene	Locus	Name
1	GGGTGGGAAAATAGACCAATAGG	38.5	1MMs [17]	NM_000518	chr11:+5248032	HBB
3	GTGTGGGAAGATGGACTAATAAG	1	3MMs [2:10:13]	NM_001013742	chrX:-50167164	DGKK
10	AGATGGTAAATTAGACTAATTGG	0.5	4MMs [1:3:7:11]	NM_001260	chr13:+26959269	CDK8
15	TGGTGGTAACATAGACTAACAGG	0.4	4MMs [1:7:10:20]	NM_012374	chr17:-56232602	OR4D1
54	GGGTAGGGAAACAGACTAGTTAG	0.2	4MMs [5:8:12:19]	NR_003529	chr9:-22097227	CDKN2B-AS1
62	GGGAGGGAACCTAGGCTAATGGG	0.2	4MMs [4:10:11:15]	NM_006377	chr9:+35398134	UNC13B
82	GGGTGTGAATACGGACTAATTGG	0.1	4MMs [6:10:12:13]	NM_004589	chr17:-10595290	SCO1
102	GGGGGGGAAAAGAGACAAAGGAG	0.1	4MMs [4:12:17:20]	NM_018088	chr12:-8378357	FAM90A1
119	GAGTGGGAAAACAGACCAGTAGG	0.1	4MMs [2:12:17:19]	NM_032878	chr19:+36503885	ALKBH6
143	GGTTGGGAAAACAGACTCAAAGG	0.1	4MMs [3:12:18:20]	NM_001177387	chr3:-63985224	ATXN7
149	TGGTGGGAAAAGAGAGTAACAGG	0.1	4MMs [1:12:16:20]	NR_040251	chr14:+74375970	ZNF410
156	GGGGGGGAAAAAAGAATAAAAGG	0.1	4MMs [4:12:16:20]	NR_003370	chr16:+21814059	RRN3P1
201	GGGTGGTAAAACAGACTCCTGAG	0	4MMs [7:12:18:19]	NM_173690	chr9:+127708667	SCAI
203	GGGAGGGAAAATAAACAATTCAG	0	4MMs [4:14:17:19]	NM_025160	chr1:+224592273	WDR26
237	GGGTGGGATAATCAACTACTAAG	0	4MMs [9:13:14:19]	NM_001029835	chr7:-45078125	CCM2

6.3 Oligonucleotide sequences / constructs / Molecular biology and cell culture reagents and equipment

Table 22: Primers and probes.
Sequences of primer and probes categorised based on their applications (Sanger sequencing, production of PCR products for T7E1 assays or sequencing, titration of LVs and VCN quantification in transduced samples by qPCR and gene expression by RT-qPCR)

Oligonucleotides for PCR/qPCR/RT-qPCR		
Primer	Application	Sequence 5' to 3'
SeqU6P FW	Sequencing	TGGATCCGGTACCAAGGTCG
Seq TAL FW	Sequencing	GCC GTG GAA GCC GTG C
Seq TAL RV	Sequencing	TCA GGG CGG CCA GAG C
NewB_RV	PCR/Sequencing	CACAGTGCAGCTCACTCAG
M13_U_FW	Sequencing	GTAAAACGACGGCCAG
M13_U_RV	Sequencing	CAGGAAACAGCTATGAC
109 RV	Sequencing/T7E1 Assay	CCCTTCCTATGACATGAACTTAACCAT
CMV FW	Sequencing	GAGACTTGGAAATCCCCGTGA
IVSII-745 FW	PCR/T7E1 Assay	ATG TAT CAT GCC TCT TTG CAC C
IVSII-745 RV	PCR/T7E1 Assay	TTG CCA AAG TGA TGG GCC AG
GCB1 FW	PCR/T7E1 Assay	TTC ACT AGC AAC CTC AAA CAG ACA CC
HBB EX2.2 RV	PCR/T7E1 Assay	CAG CTC ACT CAG TGT GGC AA
HBD EX2.1 RV	PCR/T7E1 Assay	GCA GCT CAC TCA GCT GAG AA
#985 CCR5 FW	PCR/T7E1 Assay	AAGATGGATTATCAAGTGTCAAGTCC
#986 CCR5 RV	PCR/T7E1 Assay	CAAAGTCCCACTGGGCG
Lenti_FW	Titration (qPCR)	TGAAAGCGAAAGGGAAACCA
Lenti_RV	Titration (qPCR)	TTGCCGTGCGCGCTTCAG
Lenti_Probe	Titration (qPCR)	HEX-AGCTCTCTCGACGCAGGACTCGGC-TAMRA
PCBP2_FW	Titration (qPCR)	CTGCATAATCGGGCGTCAAG
PCBP2_RV	Titration (qPCR)	GCAGCAGATCCAGTGATGGTAACCT
PCBP2_Probe	Titration (qPCR)	FAM-CGCCAAAATCAATGAGATCCGTCAGATGTCT-TAMRA
qLenti_FW	Titration (qPCR)	TCTCGACGCAGGACTCG
qLenti_RV	Titration (qPCR)	TACTGACGCTCTCGCACC
qLenti_Probe	Titration (qPCR)	YAKIMA YELLOW-ATC TCT CTC CTT CTA GCC TC-ZNA-4BHQ-1
qPCBP2_FW2	Titration (qPCR)	TTGTGTCTCCAGTCTGCTTG
qPCBP2_RV2	Titration (qPCR)	AGGTGGTGGTGGTGGTA
qPCBP2_Probe2	Titration (qPCR)	6-FAM-CCCTCTCCTGGCTCTAAATGTTGTGT-BHQ-1
GCB1 FW	Expression (RT-qPCR)	CTGAGTGAGCTGCACTGTG
GCB1 RV	Expression (RT-qPCR)	GATCCCCAAAGGACTCAAAGAACC
hHBB_EX1_FW_3	Expression (RT-qPCR)	GGGCAAGGTGAACGTG
hHBB_EX2_RV_1	Expression (RT-qPCR)	GGACAGATCCCCAAAGGAC
wtHBB Probe_ZNA	Expression (RT-qPCR)	6-FAM- TGG G(PDC)A GG(PDC) TG(PDC) TG-ZNA-3-BHQ-1
IVSI-110 MGB_Probe	Expression (RT-qPCR)	VIC-TAAGGGTGGGAAAATAGA-MGB
hHBB_EX1_RV_2_A	Expression (RT-qPCR)	CACCACCAACTTCATCCAC
hHBB_EX1_FW_1	Expression (RT-qPCR)	GGTGCATCTGACTCCTGAG
hHBB_FW_EX2_B	Expression (RT-qPCR)	GGCAAGAAAGTGCTCGG

hHBB_EX2.3_RV_B	Expression (RT-qPCR)	GTGCAGCTCACTCAGTG
mouse β-Actin Fw	Expression (RT-qPCR)	GCTTCTTTGCAGCTCCTTCGT
mouse β-Actin Fw	Expression (RT-qPCR)	CCAGCGCAGCGATATCG
m18S RNA FW	Expression (RT-qPCR)	CACGGCCGGTACAGTGAAAC
m18S RNA RV	Expression (RT-qPCR)	AGAGGAGCGAGCGACCAA
mHba-a1/a2 FW	Expression (RT-qPCR)	GTCACGGCAAGAAGGTCGC
mHba-a1/a2 RV	Expression (RT-qPCR)	GGGGTGAAATCGGCAGGG T
hHBA_FW	Expression (RT-qPCR)	GGACCCGGTCAACTTCAA
hHBA_RV	Expression (RT-qPCR)	CGGTATTTGGAGGTCAGCAC
IVS1-110 Clon_FW	PCR cloning (PCR2.1)	AAGCGGCCGCGATATCGTCACAGTGCAGCTCACTC
IVSI-110 Clon_RV	PCR cloning (PCR2.1)	AATCTAGACTCGAGTTCACTAGCAACCTCAAAC
Mut IVSI-110 FW	PCR mutagenesis	CTCTCTCTGCCTATTAGTCTATTTTCCCACCC
Mut IVSI-110 RV	PCR mutagenesis	GGGTGGGAAAATAGACTAATAGGCAGAGAGAG

Table 23: Oligonucleotides for oligo-annealing.
Upper cases sequence of interest (insert), lowercases the extension for formation of the suitable overhangs post – oligo - annealing

Oligonucleotides for Oligo-annealing		
Oligonucleotide	Application	Sequences 5' - 3'
HBB-rep-F	GFP reporter production	taaTCTTCTGATAGGCACTGACTCTCTCTGCCTATTAGTCTATTTTCCCACCCTTAgga
HBB-rep-R	GFP reporter production	ccggtccTAAGGGTGGGAAAATAGACTAATAGGCAGAGAGAGTCAGTGCCTATCAGAAGAttaat
HBB-CR-F	gRNA production	acaccGGGTGGGAAAATAGACTAATG
HBB-CR-R	gRNA production	aaaaCATTAGTCTATTTTCCCACCCG
IVSI-745 FW1	gRNA production	acaccGCTAATAGCAGCTACAATCCG
IVSI-745 RV1	gRNA production	aaaaCGGATTGTAGCTGCTATTAGCG
IVS1-745 FW2	gRNA production	acaccGATAAAAGCAGAATGGTACCG
IVS1-745 RV2	gRNA production	aaaaCGGTACCATTCTGCTTTTATCG
shIVSI-110_U_oligFW	pLKO.I_shIVSI-110	ccggGAGGCCCTGGGCAGTCTATTTGTTAACAAATAGACTGCCCAGGGCCTCTTTTTG
shIVSI-110_U_oligRV	pLKO.I_shIVSI-110	AATTCAAAAAGAGGCCCTGGGCAGTCTATTTGTTAACAAATAGACTGCCCAGGGCCTC
shIVSI-110_M_oligFW	pLKO.I_shIVSI-110	ccggGTCTATTTTCCCACCCTTAGGAGTACTCCTAAGGGTGGGAAAATAGACTTTTTG
shIVSI-110_M_oligRV	pLKO.I_shIVSI-110	AATTCAAAAAGTCTATTTTCCCACCCTTAGGAGTACTCCTAAGGGTGGGAAAATAGAC
shIVSI-110_M2_oligFW	pLKO.I_shIVSI-110	ccggGTCTATTGTCCCACCCTTAGGAGTACTCCTAAGGGTGGGAAAATAGACTTTTTG
shIVSI-110_M2_oligRV	pLKO.I_shIVSI-110	AATTCAAAAAGTCTATTTTCCCACCCTTAGGAGTACTCCTAAGGGTGGGACAATAGAC
shIVSI-110_D_oligFW	pLKO.I_shIVSI-110	ccggTTTCCCACCCTTAGGCTGCTGAGTACTCAGCAGCCTAAGGGTGGGAAATTTTTG
shIVSI-110_D_oligRV	pLKO.I_shIVSI-110	AATTCAAAAATTTCCCACCCTTAGGCTGCTGAGTACTCAGCAGCCTAAGGGTGGGAAA
shIVSI-110_D2_oligFW	pLKO.I_shIVSI-110	ccggTTTCCCACCCTTAGGCTGCTGTTCAAGAGACAGCAGCCTAAGGGTGGGAAATTTTTG
shIVSI-110_D2_oligRV	pLKO.I_shIVSI-110	AATTCAAAAATTTCCCACCCTTAGGCTGCTGTCTCTTGAACAGCAGCCTAAGGGTGGGAAA

Table 24: Plasmids.

Plasmids	Size (bp)	Description
#277.pCCLsin.PPT.hPGK.GFP.pre	7827	LV transfer plasmid: eGFP transgene under the human PGK promoter
MA_821 (GLOBE)	9971	LV transfer plasmid: mini LCR(HS2/3) and β-globin promoter, driving a mini human β-globin gene (deletion in intron 2)
MA_821_shBCL11A_shGFP	9924	LV transfer plasmid: T87Q with shBCL11A and shGFP in intron 2
MA_821_T87Q	9971	LV transfer plasmid: MA821 with T87Q substitution in human β-globin
MA_821_T87Q_shIVS1-110_Down	9844	LV transfer plasmid: T87Q with shIVS1-110_Down in intron 2
MA_821_T87Q_shIVS1-110_Mid	9845	LV transfer plasmid: T87Q with shIVS1-110_Mid in intron 2
MA_821_T87Q_shIVS1-110_Up	9855	LV transfer plasmid: T87Q with shIVS1-110_Up in intron 2
pADVantagepADvantage	4392	Encoded adeno associated virus RNA I , II. EnchancesEnhances transient protein expression in a variety of cell types by increasing translation initiation
pEGFP_N1	4733	TranfectionTransfection Reporter plasmid: eGFP transgene under CMV promoter
pCMVdR8.74	11875	Packaging plasmid: expressing polymerase/capsid proteins
pCR2.1_HBB_cDNA normal	4238	Construct cloned with fragment of correctly spliced HBB cDNA. Use as SC for assessment of correctly spliced HBB mRNA
pCR2.1_HBB_cDNA ^{Aberrant}	4257	Construct cloned with fragment of abberantlyaberrantly spliced HBB cDNA. Use as SC for assessment of aberrantly spliced HBB mRNA
pCR2.1_HBB_cDNA ^{Normal}_cDNA^{Aberrant}	4550	Construct cloned with fragment of correctly and aberrantly spliced HBB cDNA. Use as SC for accurate assessment of both HBB mRNA variants
pJDS246 (Cas9)	7614	spCas9 Plasmid: expessionexpression of S. Pyogenes Cas9 under CMV promoter
pLKO.I shIVSI-110_Down	7084	LV transfer plasmid: Expressing of shIVSI110_Down under the U6 promoter
pLKO.I shIVSI-110_Middle	7084	LV transfer plasmid: Expressing of shIVSI110_Middle under the U6 promoter
pLKO.I shIVSI-110_Middle_2	7084	LV transfer plasmid: Expressing of shIVSI110_Middle_2 under the U6 promoter
pLKO.I shIVSI-110_Scramble	7085	LV transfer plasmid: Expressing of shScramble under the U6 promoter
pLKO.I shIVSI-110_Up	7084	LV transfer plasmid: Expressing of shIVSI110_Up under the U6 promoter
pLML3636 gRNA vector	2778	pLML3636 gRNA plasmid: gRNA Vector under U6 promoter (cloning site *BsiB*I)
pLML3636 HBB $^{IVSI-110}$ gRNA	2779	HBB $^{IVSI-110}$ gRNA plasmid: HBB$^{IVSI-110}$ specific gRNA construct under U6 promoter
pLML3636 HBB$^{IVSII-745}$ A gRNA	2779	HBB IVSII-745 gRNA A plasmid:VSII-745 specific gRNA A gRNA construct under U6 promoter
pLML3636 HBB$^{IVSII-745}$ B gRNA	2779	HBB IVSII-745 gRNA B plasmid:VSII-745 specific gRNA B gRNA construct under U6 promoter
pLSSmOrangeN1	4723	TranfectionTransfection Reporter plasmid: mOrange transgene under CMV promoter
pLV.CMV HBB$^{IVSI-110}$ dsEGFP HBB reporter	9067	Targeted disruption HBB$^{IVSI-110}$ eGFP reporter plasmid: HBB$^{IVSI-110}$ target sequence in-frame upstream from the dsEGFP cDNA sequence under CMV promoter
pLV.CMV.dsEGFP	9064	Target Disruption reporter Vector: ClonningCloning site (PacI/AgeI) between ATG and dseGFP trangenetransgene under CMV promoter
pLV.CMV.HBB $^{IVSI-}$	9067	Target Disruption reporter plasmid: HBB $^{IVSI-110}$ target sequence between

[110] dsEGFP		ATG and dseGFP trangenetransgene under CMV promoter
pMD2.VSVG	5824	Envelope plasmid: expressing vesicular stomatitis virus envelope glycoprotein (VSVG)
pMK-RQ_sh_IVS1-110_Down	2360	Synthetic construct with shIVS1-110_Down insert (*GeneArt*)
pMK-RQ_sh_IVS1-110_Middle	2361	Synthetic construct with shIVS1-110_Mid insert (*GeneArt*)
pMK-RQ_sh_IVS1-110_Up	2361	Synthetic construct with shIVS1-110_Up insert (*GeneArt*)
pRK5.mCherry	5424	Transfection Reporter plasmid: mCherry transgene under CMV promoter [228]
pRSV-Rev	4174	Accessory REV plasmid: encodes REV
pUC118 pVAX_CMV_HBB_TAL_L1_FokI	3162 5885	pUC derivative. Used to normalize for the amount of plasmid DNA. in transfections or serve the nuclease Free negative control HBB TALEN monomer L1:expression of the HBB specific L1 monomer
pVAX_CMV_HBB_TAL_L2_FokI	5885	HBB TALEN monomer L1:expression of the HBB specific L2 monomer
pVAX_CMV_HBB_TAL_R1_FokI	5885	HBB TALEN monomer L1:expression of the HBB specific R2 monomer (0 NK 6 NN RVG)
pVAX_CMV_HBB_TAL_R2_FokI	5885	HBB TALEN monomer L1:expression of the HBB specific R2 monomer (2 NK 4 NN RVG)
pVAX_CMV_HBB_TAL_R3_FokI	5885	HBB TALEN monomer L1:expression of the HBB specific R3 monomer (4 NK 2 NN RVG)
pVAX_CMV_HBB_TAL_R4_FokI	5885	HBB TALEN monomer L1:expression of the HBB specific R4 monomer (6 NK 0 NN RVG)

Table 25: Molecular biology reagents and equipment

Molecular Biology reagents/equipment	Catalog number	Company
1 kb DNA ladder	N3232S	New England Biolabs Inc. (Massachusetts, USA)
100 mM dGTP, dATP, dCTP, dTTP	473876, 473877, 406883, 473875	Life Technologies (California, USA)
250 µL microinserts	2109073	GraceVydac, USA
3. 2-log DNA ladder (0.1 – 10kb)	N3200S	New England Biolabs Inc. (Massachusetts, USA)
40% Acrylamide/Bis Solution, 37.5:1	161-0148	Bio-Rad Laboratories, Inc. (California, USA)
ABI 3130xl Genetic Analyzer		Life Technologies (California, USA)
Acetic acid	1000632500	Merck Millipore (Massachusetts, US)
Acetonitrile	34888	Sigma-Aldrich (Missouri, USA)
Agar, granulated	GM1002	Melford (Ipswich, UK)
Agarose	A9539-500G	Sigma-Aldrich (Missouri, USA)
Albumin, from bovine serum	A3912-50G	Sigma-Aldrich (Missouri, USA)
Ammonium peroxodisulfate (APS)	1.01201.0100	Merck KGaA (Darmstadt, Germany)
Ampicillin	A9393-5G	Sigma-Aldrich (Missouri, USA)
Anti-Ctip1 antibody [15E3AC11]	ab18688	Abcam plc (Cambridge, UK)
BigDye Terminator v3.1 Cycle Sequencing kit	4337455	Life Technologies (California, USA)
Bio-Rad Imagine System		Bio-Rad Laboratories, Inc. (California, USA)

BlueStar Prestained Protein Marker	MWP03	NIPPON Genetics (Duren, Germany)
Bromophenol Blue (BFB)	B0126	Sigma-Aldrich (Missouri, USA)
complete, EDTA-free (protease inhibitor)	11873580001	Roche (Basel, Switzerland)
Criterion™ Blotter With Plate Electrodes	170-4070	Life Technologies (California, USA)
CyFlow® Cube 6	CY-S-3060_V1_S	Sysmex Co. (Kobe, Japan)
DNase I Amplification Grade	18068-015	*Invitrogen™, Thermo* Fisher *Scientific*
Dual Labeled ZNA probe		*Metabion International AG*, Planegg, Germany
Ethanol	1.00983.2500	Merck KGaA (Darmstadt, Germany
Ethylenediaminetetraacetic acid (EDTA)	8.19040.0100	Merck Millipore (Massachusetts, US)
EXOSAP-IT	78201	Affymetrix, Thermo Fisher Scientific
Gene Pulser® II RF Module	165-2113	Bio-Rad Laboratories, Inc. (California, USA)
Glycerol	101186M	AnalaR, BDH (Pennsylvania, USA)
Glycine	G/0800/60	Fisher Scientific International, Inc (New Hampshire,USA)
H_2O: Merck LiChrosolv	1.15333.2500	Merck Millipore, USA
HPLC vials	2109299	GraceVydac, USA
Hydrochloric acid (HCl)	101256J	AnalaR, BDH (Pennsylvania, USA)
Hydrogen peroxide solution 30%	1.08597.1000	Merck KGaA (Darmstadt, Germany)
ImageLab Software 5.1	1709690	Bio - Rad
Jupiter 5 µm C18 25 cm 4.6 mm	00G-4053-E0	Phenomenex
kanamycin	17924 5 GM	USB, Affymetrix, Thermo Fisher Scientific
LumiSensor Chemiluminescent HRP Substrate kit	L00221V300	GenScript (New Jersey, USA)
Luria Broth	12795-027	Life Technologies (California, USA)
Methanol	32213-2.5L	Sigma-Aldrich (Missouri, USA)
Methanol: Merck LiChrosolv	1.06018.2500	Merck Millipore, USA
MicroAmp Optical 96-well Reaction Plate	N8010560	Life Technologies (California, USA)
MicroPulser Electroporation Cuvettes (0.4 cm)	1652081	Bio-Rad Laboratories, Inc. (California, USA)
Millex-GV Syringe Filter Unit, 0.22 µm	SLGV033RS	Merck Millipore (Massachusetts, USA)
Millex-HV 0.45-μm pore size PVDF membrane filter		*Millipore, Darmstadt, Germany*
Mini-PROTEAN Tetra Cell	165-8001EDU	Bio-Rad Laboratories, Inc. (California, USA)
Mouse monoclonal anti - human HBB (37-8)	sc-21757	Santa Cruz biotechnologies
Mouse monoclonal anti – murine Actb antibody (AC-15)	A1978	Sigma-Aldrich (Missouri, USA)
mouse monoclonal HA-probe Antibody (F-7)	sc-7392	Santa Cruz Biotechnologies
ND-1000 Spectrophotometer		NanoDrop Technologies, Inc. (Delaware, US)

Nonidet™ P 40 Substitute (NP40)	74385	Sigma-Aldrich (Missouri, USA)
NucleoBond Xtra Midi and Maxi	740410.10 and 740414.10	MACHEREY-NAGEL (Duren, Germany)
NucleoSpin Gel and PCR Clean-up	740609.25	MACHEREY-NAGEL (Duren, Germany)
Orange G	3756	Sigma-Aldrich (Missouri, USA)
PERFORMA DTR Gel Filtration Cartridges	42453	EdgeBio (Maryland, USA)
Peroxidase-conjugated AffiniPure goat anti-mouse IgG (H=L)	115-035-003	Jackson ImmunoResearch Laboratories, PA, USA
Peroxidase-conjugated AffiniPure goat anti-rabbit IgG (H=L)	111-035-003	Jackson ImmunoResearch Laboratories, PA, USA
Phusion High-Fidelity DNA polymerase	M0503L	New England Biolabs Inc. (Massachusetts, USA)
Ponceau S solution	81462	Sigma-Aldrich
Porablot NCP Nitrocellulose membrane	741280	MACHEREY-NAGEL (Duren, Germany)
Potassium Acetate	P1190	Sigma-Aldrich (Missouri, USA)
Propan-2-ol (isopropanol)	24137-2.5L-R	Sigma-Aldrich (Missouri, USA)
Protease	1017784	QIAGEN N.V (Hilden, Germany)
QIAamp DNA Blood Mini kit	51106	QIAGEN N.V (Hilden, Germany)
Qiagen Multiplex PCR Kit	206143	QIAGEN N.V (Hilden, Germany)
QIAquick PCR purification kit	28106	QIAGEN N.V (Hilden, Germany)
QuikChange Lightning Site-Directed Mutagenesis Kit	210518	*Agilent Technologies, USA*
Rabbit polyclonal anti - murine alpha (H-80)	sc-21005	Santa Gruz biotechnologies
Rapid DNA Ligation kit	11 635 379 001	Roche (Basel, Switzerland)
Ribonuclease A from bovine pancreas	R4875-100MG	Sigma-Aldrich (Missouri, USA)
RNase A	19101	QIAGEN N.V (Hilden, Germany)
Shimadzu LC-20AD chromatographic system		*Shimadzu, Kyoto, Japan*
Sodium Chloride (NaCl)	31434-1KG-R	Sigma-Aldrich (Missouri, USA)
Sodium deoxycholate	D6750	Sigma-Aldrich (Missouri, USA)
Sodium Dodecyl Sulfate (SDS) 20%	BP1311-1	Fisher Scientific International, Inc (New Hampshire, USA)
Sodium Hydroxide (NaOH)	06203-1KG	Sigma-Aldrich (Missouri, USA)
SYBR Green PCR Master Mix	4309155	Life Technologies (California, USA)
SYTOX Red Dead Cell Stain	S34859	Life Technologies (California,USA)
T7 Endonuclease 1 (T7E1)	M0302L	New England Biolabs Inc. (Massachusetts, USA)
TA Cloning® Kit with One Shot® TOP10 Chemically Competent E. coli	K2040-01	Thermo Fisher Scientific
TaqMan PCR Master mix	S04021	Applied Biosystem , Thermo Fisher Scientific
TaqMan Reverse Transcription Reagents	N808-0234	Life Technologies (California, USA)
TaqMan® Reverse Transcription Reagents kit	N8080234	Applied Biosystem , Thermo Fisher Scientific
Tetramethylethylenediamine(TEMED)	1.10732.0100	Merck KGaA (Darmstadt, Germany)
Trifluoroacetic acid (TFA)	73645	Sigma-Aldrich (Missouri, USA)
Trizol	15596026	*Invitrogen™*, Thermo Fisher Scientific

Turbo Competent Escherichia coli (High Efficiency)	C2984H	New England Biolabs Inc. (Massachusetts, USA)
Veriti 96-well Thermal Cycler	4375786	Life Technologies (California, USA)
VIC-MGB Probe	4316034	Life Technologies (California, USA)
ZERO TOPO CLONE KIT SEQU with One Shot® TOP10 Chemically Competent E. coli	K287520	Thermo Fisher Scientific

Table 26: Cell culture reagents and equipment

Cell culture reagents/equipment	Catalog number	Company
145 cm2 Cell Culture Petri Dish With Lid, Vented, Nunclon™ Delta Surface	168381	Thermo Fisher Scientific (Massachusetts, USA)
24-well Nunc non-treated multidishes	144530	Thermo Fisher Scientific (Massachusetts, USA)
25 cm^2 Nunclon™ Delta flasks	156367	Thermo Fisher Scientific
6-well plate Nunc non-treated multidishes	140685	Nunc; Thermo Fisher Scientific
75 cm^2 Rectangular Canted Neck Cell Culture Flask	430641	Corning Inc. (New York, USA)
80 cm^2 Cell Culture Flask, Nunclon™ Delta Surface	178905	Thermo Fisher Scientific (Massachusetts, USA)
anti-hCD34$^+$ MicroBeads		Miltenyi Biotecy, Bergisch Gladbach, Germany
CC100	2690	Stemcell Technologies, Vancouver, Canada
Costar 6-well clear tissue culture-treated multiple well plates	3516	Corning Inc. (New York, USA)
dexamethasome	D4902	Sigma-Aldrich
Dimethyl sulfoxide (DMSO)	D8428-250ML	Sigma-Aldrich (Missouri, USA)
Dulbecco's Modified Eagle Medium (DMEM)	41965-039	Life Technologies (California, USA)
Dulbecco's Phosphate Buffered Saline (DPBS) x1	14190-094	Life Technologies (California, USA)
EDTA-coated tubes	455036	Greiner Bio-One, Frickenhausen, Germany
erythropoietin (EPO) 1000 U/mL		Sandoz GmbH, Austria, Binocrit
FetalFoetal Bovine Serum (Characterized) Hyclone	SH30071.03	Thermo Fisher Scientific
FetalFoetal Bovine Serum (Define) Hyclone	SH30070.03	Thermo Fisher Scientific
Giemsa's stain		Fluka Analytica, Sigma-Aldrich, UK)
GlutaMAX™-I x100	35050-038	Life Technologies (California, USA)
LS MACS columns		Miltenyi Biotecy, Bergisch Gladbach, Germany
Lymphoprep -Accu-Prep™ Lymphocytes		Axis-Shield PoC AS, Norway
May-Grünwald stain		Sigma-Aldrich, UK
MidiMACS separator and Multistand		Miltenyi Biotecy, Bergisch Gladbach, Germany
Mounting medium		Dakocytomation, CA, USA
NuaireDH Auto Flow CO_2 Air Jacket incubator		Polymouth, USA
o-Dianisidine		Sigma-Aldrich
Penicillin/Streptomycin	30-002-CI	Corning-CellGro
Penicillin-Streptomycin (10,000 U/ml)	15140-122	Life Technologies (California,

		USA)
Petri Dish 94/15MM	633179	Greiner Bio-one (Kremsmünster, Upper Austria)
Polyethylenimine (PEI), linear, MW 25000Da	23966	Polysciences Inc. (Pennsylvania, USA)
RPMI Medium 1640 x1	21875-034	Life Technologies (California, USA)
Stem Cell Factor (SCF)	300-07	Peprotech
StemSpam II	9650	Stemcell Technologies, Vancouver, Canada
Tharmac Cellspin II cytocentrifuge EASY rotor	A320	Hettich, Germany
αMEM	10-22-CV	Corning CellGro
10^{-5}M 2-mercaptoethanol	M3148	Sigma-Aldrich
Trypan Blue Solution (0.4%)	T18154	Sigma-Aldrich (Missouri, USA)
0.25% Trypsin-EDTA x1	25200-072	Life Technologies (California, USA)

6.4 TALENs and RGENs sequences

Table 27: TALEN's RVD sequence.
The [T] at 0 position are excluded from RVD sequence.

TALEN monomers	**RVD sequences N - terminal ' – C - terminal'**	**Binding sequence (5' to 3') on the target gene**
HBB TAL R1	NI NI NN NN NN NG NN NN NN NI NI NI NI NG NI NN NI HD	TAAGGGTGGGAAAATAGAC
HBB TAL R2	NI NI NK NK NN NG NN NN NN NI NI NI NI NG NI NN NI HD	TAAGGGTGGGAAAATAGAC
HBB TAL R3	NI NI NK NK NK NG NK NN NN NI NI NI NI NG NI NN NI HD	TAAGGGTGGGAAAATAGAC
HBB TAL R4	NI NI NK NK NK NG NK NK NK NI NI NI NI NG NI NN NI HD	TAAGGGTGGGAAAATAGAC
HBB TAL L1	NG HD NG NN NI NG NI NN NN HD NI HD NG NN NI HD NG HD	TTCTGATAGGCACTGACTC
HBB TAL L2	NN NI NG NI NN NN HD NI HD NG NN NI HD NG HD NG HD NG	TGATAGGCACTGACTCTCT
RM98 CCR5 TAL L [227]	NG NG NN NG NN NN NN HD NI NI HD NI NG NN HD NG NN NN	TTTGTGGGCAACATGCTGG
RM101 CCR5 TAL R [227]	HD NI NN HD HD NG NG NG NG NN HD NI NN NG NG NG NI NG	TCAGCCTTTTGCAGTTTAT

Table 28: gRNA sequences.

gRNAs	**gRNA 5' - 3'NGG PAM in parenthesis**	**Orientation on target gene**
IVSI-110 gRNA	5'-GGGTGGGAAAATAGACT**A**AT (AGG) -3'	- strand
IVSII-745 gRNA A	5'-GCTAATAGCAGCTACAATCC (AG**G**)-3'	+ strand
IVSII-745 gRNA B	5'-AATAAAAGCAGAATGGTA**C**C (TGG)-3'	- strand
CCR5 gRNA	5'- GTGAGTAGAGCGGAGGCAGG (AGG)	- strand

6.5 Publications

OPEN

Gene Therapy (2016) 23, 113–118

www.nature.com/gt

SHORT COMMUNICATION

Measurement of lentiviral vector titre and copy number by cross-species duplex quantitative PCR

I Christodoulou[1], P Patsali[2,3], C Stephanou[2,3], M Antoniou[3], M Kleanthous[1,2,4] and CW Lederer[1,2,4]

Lentiviruses are the vectors of choice for many preclinical studies and clinical applications of gene therapy. Accurate measurement of biological vector titre before treatment is a prerequisite for vector dosing, and the calculation of vector integration sites per cell after treatment is as critical to the characterisation of modified cell products as it is to long-term follow-up and the assessment of risk and therapeutic efficiency in patients. These analyses are typically based on quantitative real-time PCR (qPCR), but as yet compromise accuracy and comparability between laboratories and experimental systems, the former by using separate simplex reactions for the detection of endogene and lentiviral sequences and the latter by designing different PCR assays for analyses in human cells and animal disease models. In this study, we validate in human and murine cells a qPCR system for the single-tube assessment of lentiviral vector copy numbers that is suitable for analyses in at least 33 different mammalian species, including human and other primates, mouse, pig, cat and domestic ruminants. The established assay combines the accuracy of single-tube quantitation by duplex qPCR with the convenience of one-off assay optimisation for cross-species analyses and with the direct comparability of lentiviral transduction efficiencies in different species.

Gene Therapy (2016) **23,** 113–118; doi:10.1038/gt.2015.60; published online 23 July 2015

INTRODUCTION

Lentiviral vectors (LVs) allow curative gene therapy (GT) through genome integration, accept large payloads at high vector titre, readily transduce non-dividing cells, have a favourable safety profile compared with other integrating viral vectors and, depending on pseudotype, have wide tissue and host tropism, and thus wide-ranging applicability in the field of GT.[1] Efficacy and safety of their application are critically determined by the average vector copy number (VCN) in the corrected cell product, with more integration sites generally giving higher transgene expression levels in gene augmentation approaches, while also increasing the number of viral integration sites and thus the risk of insertional mutagenesis.[2] The desired VCN, as dictated by the required expression levels and percentage of corrected cells on the one hand and by the efficacy, long-term stability and homogeneity of transgene-derived gene expression on the other, is approximated by application of a specific number of biologically active vector particles to a known number of target cells. The actual VCN in the cell product in combination with culture and transplantation regimens, and with a possible growth advantage of corrected cells in turn determine the average VCN in the target tissues of treated patients post GT.

Initial biological vector titre, actual VCN in the cell product and post-GT VCN in the patient are usually determined by the same quantitative real-time PCR (qPCR) procedure, optimised for the respective system under study.[3] The principle of VCN determination for integrating vectors is the quantification of proviral vector genomes in extracted genomic DNA (gDNA) and their normalisation by the number of haploid host genomes against a standard curve of plasmid dilutions or of transgenic material of known copy number. The choice of LV and in particular of endogene target sequences differs considerably between laboratories, but the underlying qPCR assays almost invariably rely on (for example, TaqMan, Applied Biosystems, Waltham, MA, USA) probe technology instead of double-stranded DNA-integrating dyes, in order to minimise the detection of spurious byproducts that is common to the latter detection method. The application of labelled probes in turn opens up the possibility of duplexing LV- and endogene-specific qPCR assays labelled with different detection dyes, in order to allow same-tube detection of both sequences and with the inherent benefits of removing same-sample tube-to-tube variation and reducing reagent cost and processing time. However, to date only the Kohn group has developed a duplex protocol for LV VCN determination,[4] which has as yet not been used independently.[5–8] Strikingly and although any GT clinical trial is as a rule preceded by preclinical validation in non-human disease models, only one assay, incompatible with duplex application, has as yet been published that uses the same endogene qPCR assay for human and animal LV-GT samples.[4] Application of independent qPCR assays, however, precludes a direct comparison of VCN quantitation between species and, on the practical side, duplicates the effort and cost of assay optimisation. As an overriding dilemma, the absence of a community-wide standard quantitation assay and the habitual underdocumentation of technical details pertaining to VCN determination in outcome-focussed publications compromise the comparability of VCN and vector activity per copy between different research groups. This study therefore set out to establish

[1]Cyprus School of Molecular Medicine, Nicosia, Cyprus; [2]Department of Molecular Genetics Thalassaemia, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus and [3]King's College London, Gene Expression and Therapy Group London, UK. Correspondence: Dr CW Lederer, Department of Molecular Genetics Thalassaemia, The Cyprus Institute of Neurology and Genetics, 6 International Airport Avenue, Nicosia 1683, Cyprus.
E-mail: Lederer@cing.ac.cy
[4]These authors contributed equally to this work.
Received 17 October 2014; revised 1 April 2015; accepted 6 May 2015; published online 23 July 2015

a duplex, cross-species assay for LV VCN quantitation towards the establishment of a community standard for VCN reporting.

RESULTS AND DISCUSSION

Aiming to establish a qPCR assay with cross-species compatibility for human and murine sequences, we chose a published LV qPCR assay,[9] qLV (see Table 1), and selected the highly conserved poly (rC)-binding protein 2 (*PCBP2*) gene from the literature,[10] and a *PCBP2* region of high sequence identity and sufficient length for optimised endogene primer and probe design. The corresponding qPCR assay, qPCBP2, has perfect sequence identity of primers and probes for 35 different mammalian species (see Figure 1a), including human, apes, new- and old-world monkeys, Chinese and golden hamster, mouse, pig, domestic and wild ruminants, delphinidae, three bat species, hedgehog, shrew and armadillo, and sequence identity of the entire amplicon for 33 of these. In the following, we validate the VCN assay in human and murine cells and demonstrate the functionality of the qPCBP2 assay in three additional species, but all of the above species are either naturally or experimentally infected with lentiviruses (see, for example,[11]) and thus possible assay targets.

Duplex PCR is subject to competition of all same-tube amplicons, which for quantitative assays entails the confirmation of linear amplification for all products of interest within the prospective range and ratios of target concentrations. Following a simple workflow (see Supplementary Materials and Methods 1), we used gDNA from human and murine erythroleukaemia cell lines (HEL and MEL, respectively) to work towards a standard assay for 100 ng gDNA of lentivirally transduced sample. To this end, we tested the amplification of simplex versus duplex reactions for combinations of 50–200 ng gDNA, in order to allow for limiting sample amounts or considerable imprecision in spectrophotometrically determined DNA quantities, with lentiviral plasmid DNA (see Supplementary Materials and Methods 2). We used LV plasmid amounts equivalent to VCN 0.16 up to 20 in 100 ng of gDNA, adjusting plasmid amounts for the respective genome size and arguing that a VCN outside this range would not be used for titre calculations, and that for therapeutic applications a lower VCN would be insufficient and a higher VCN would carry an unacceptably high risk of insertional mutagenesis.

Initial attempts to establish a duplex protocol based on the kit routinely used in our laboratory for simplex detections (TaqMan 2× PCR Master Mix, Applied Biosystems) led to a depression of amplicon detection in duplex compared with equivalent simplex reactions when the alternative template became more abundant within physiological and routine experimental parameters (VCN ⩾1.6 or gDNA ⩾100 ng; data not shown). This nonlinear behaviour of duplex reactions persisted despite modification of a multitude of parameters relevant to qPCR efficiency and amplicon competition (see Materials and Methods), but disappeared after a switch to the 2× Multiplex PCR Master Mix (Qiagen, Hilden, Germany).

The dedicated multiplex kit in combination with qLV and qPCBP2 allowed us to established a standardised duplex protocol (see Supplementary Materials and Methods 3 and 4) with extreme correspondence between expected and observed values for LV detection (Supplementary Figure S1) and gDNA detection (Supplementary Figure S2), and between simplex and duplex detection of LV (Figure 2a and Supplementary Figure S3) and gDNA (Figure 2b) within the above parameters. The qPCR duplex assay displayed average amplification efficiencies of (96.8 ± 3.2)% for qLV and of (93.3 ± 4.8)% for qPCBP2 across all combinations of VCNs and gDNA template types and concentrations tested, as shown in Supplementary Table S1, which also gives detailed results for efficiencies and quality of fit for individual combinations.

Finally, and using gDNA from clones of MEL and HEL cells transduced with LVs typical for GT applications, we ascertained the consistency of the measured number of integrated proviral copies with those obtained through an independent simplex assay (see Figures 1b–e and Supplementary Figure S4). Subsequent tests of the assay with gDNA from additional species (see Supplementary Figure S5, demonstrating use with bovine, caprine and ovine gDNA) confirm its potential applicability to a wide range of model systems. The duplex VCN assay has since been adopted in our laboratory for the determination of LV titres after vector production and of VCN after transduction of human peripheral blood-derived $CD34^+$ cells and murine bone marrow-derived lineage-negative cells, and has proven robust and economical. Besides these benefits in its local application, the wide species compatibility of this assay allows its wider adoption and thus its establishment as a standard for comparing derived experimental and vector parameters, such as transduction efficiency and activity per vector copy, between independent lentiviral GT studies.

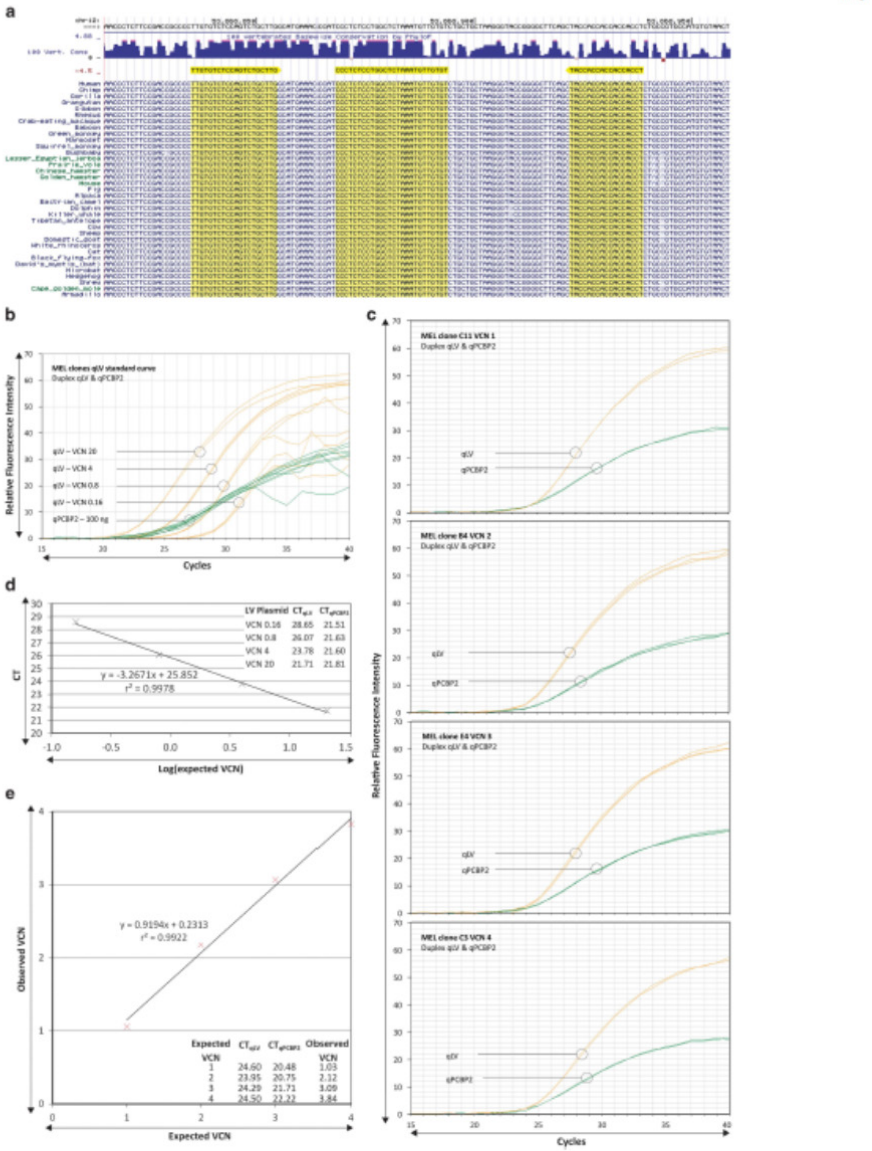
MATERIALS AND METHODS

Cell material for DNA extraction was wild-type and LV-transduced HEL (92.1.7 (ATCC TIB-180)) and MEL (48 (ATCC, Manassas, VA, USA, CRL-1913)) cell populations and clones, maintained at 37 °C, 5% CO_2 and 100% relative humidity in RPMI medium, supplemented with 1% penicillin/streptomycin, 10% fetal bovine serum and 1% L-glutamate (all from Life Technologies, Carlsbad, CA, USA). The LVs used for transduction were $MA821^{T87Q}$, a derivative of the GLOBE vector[12] encoding a T87Q mutation in the β-globin open-reading frame, and pCCL.sin.cPPT.hPGK.GFP.WPRE[13] for MEL and human erythroleukaemia cells, respectively. Cell clones were isolated by twofold limiting dilution and visual inspection of single-cell status, followed by simplex qPCR quantification using LV plasmid and gDNA standard curves. The clones used here

Table 1. Primer and probe sequences for the duplex cross-species VCN qPCR assay

qPCR assay	*Target*	*Function*	*Sequence (5′ to 3′)*
qLV	HIV gag	FW	TCTCGACGCAGGACTCG
		RV	TACTGACGCTCTCGCACC
		Probe	*Yakima-Yellow*-ATCTCTCTCCTTCTAGCCTC-*ZNA*$_4$-*BHQ1*
qPCBP2	*PCBP2* IVS 13	FW	TTGTGTCTCCAGTCTGCTTG
		RV	AGGTGGTGGTGGTGGTA
		Probe	FAM-CCCTCTCCTGGCTCTAAATGTTGTGT-BHQ1

Abbreviations: BHQ1, black hole quencher-1; FAM, 6-carboxyfluorescein; FW, forward primer; IVS, intervening sequence; probe, 5′ fluorescently labelled and 3′ quencher-linked probe; RV, reverse primer; ZNA, zip nucleic acid cationic spermine residues. Assay qLV is based on sequences published by Amendola *et al*,[9] with modifications shown in italics, and recognises the gag sequence of the $MA821^{T87Q}$ transfer vector plasmid and of proviral integration sites. Assay qPCBP2 recognises an intronic *PCBP2* region of perfect cross-species identity between mouse, human and 31 additional species for the entire amplicon.



showed a consistent quantity ratio for the LV transgene in three independent experiments. gDNA was extracted using either a standard phenol:chloroform:isopropanol–chloroform extraction method (for bovine, caprine and ovine tissues) and the FlexiGene DNA kit (Qiagen, for cell lines). The LV plasmid standard used was the corresponding 9771-bp plasmid encoding the MA821^{T87Q} transfer vector segment, with 3.26 and 3.74 pg plasmid DNA representing a VCN of 1 in 100 ng of human (3.1 Gb per haploid genome, GRCh37) and murine (2.7 Gb per haploid genome, GRCm38) gDNA, respectively. Experimental procedures followed for vector production in HEK 293T cells, vector titration and lentiviral transduction of cells have been described elsewhere.[12,14] For LV standard curves, fivefold serial dilutions of 6.3 and 7.2 pg plasmid per 25-µl reaction were used in combination with human and murine gDNA, respectively. The curves thus represented VCN equivalents of 20, 4, 0.8 and 0.16 for 100 ng gDNA, with circular

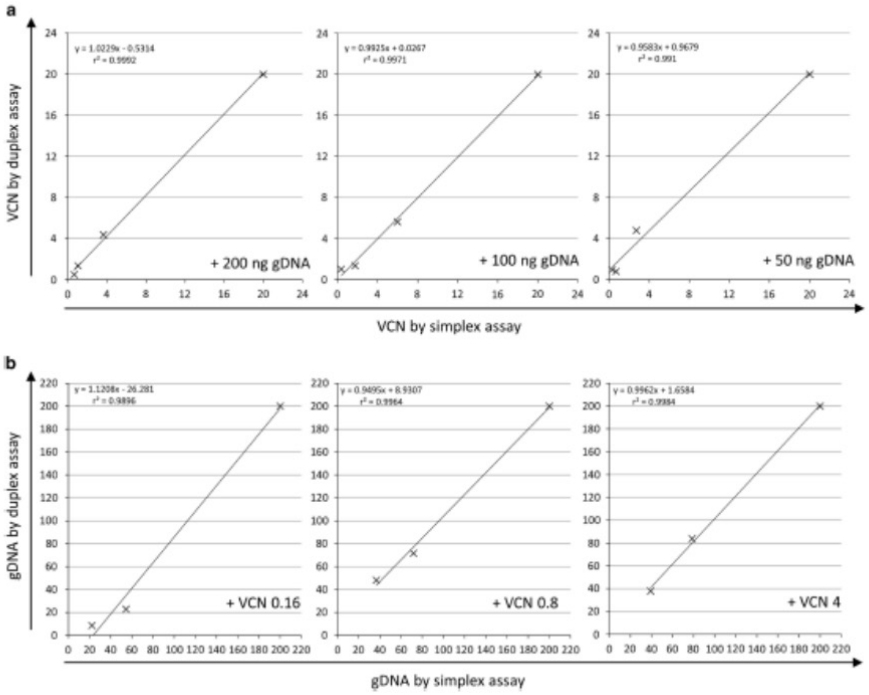


Figure 2. Simplex versus duplex quantification of LV plasmid and gDNA. The scatter plots show the correlation of template amounts determined by simplex assay with those determined by duplex assay in the presence of varying amounts of the respective alternative template, giving the formula of the trend line and the correlation (r^2) between simplex and duplex values for each plot. (**a**) Quantification of LV plasmid DNA (VCN equivalence) in the presence of varying amounts (200, 100 and 50 ng) of murine gDNA. (**b**) Quantification of murine gDNA (ng) in the presence of varying LV plasmid DNA amounts (equivalent to VCN 0.16, 0.8 and 4).

Figure 1. Sequence alignment of qPCBP2 and validation of qLV and qPCBP2 qPCR assays. (**a**) Cross-species alignment for target sequences of the *PCBP2*-specific qPCR assay *PCBP2* gene sequences were accessed using the UCSC Genome Browser (http://genome.ucsc.edu; Build GRCh37/hg19) and its Vertebrate Multiz Alignment track. Perfect sequence matches for *PCBP2* probe and primer sequences are highlighted. (**b–d**) Duplex standard curves for qLV and qPCBP2, and same-plate amplification of MEL clones with a known VCN of 1, 2, 3 and 4, respectively, for the MA821^{T87Q} LV are shown and analysed. For scatter plots, corresponding trend line formulas and curve fit values (r^2) are shown. All outlier-removed reactions of the duplex assay are shown. (**b**) Fluorescence plot of duplex qPCR for LV plasmid standard curve in the presence of 100 ng MEL gDNA, used for the absolute quantification of LV integration sites in MEL clones. The cause of plateau-phase noise is unknown. (**c**) Fluorescence plot of duplex qPCR for MEL clones with known numbers of viral integration sites. (**d**) Scatter plot of the standard curve shown in **b**. The plot shows the logarithm of expected VCNs against the automatically determined threshold cycle (CT) values and provides average CT values for each data point for qLV and qPCBP2 detection as a table insert. (**e**) Scatter plot for VCN determination of MEL clones analysed in **c**. The plot shows the expected VCN against that determined (observed) using the standard curve in **d**. The expected and observed VCN values and average CT values are additionally shown as a table insert.

and linearised plasmid (using *ScaI*-HF (New England Biolabs, Inc., Ipswich, MA, USA) for linearisation) performing equivalently. *PCBP2* was detected as an endogene in defined amounts of gDNA, both for transgenic samples and for standard curves, which comprised serial dilutions to 200, 100 and 50 ng of wild-type DNA per 25-µl reaction, either neat or laced with the MA821^{T87Q} transfer vector plasmid at the VCN equivalents given above. Quantification of viral DNA and gDNA in simplex reactions was initially performed with the TaqMan 2× PCR Master Mix (Applied Biosystems). Of the quantitative PCR assays (all from Metabion International AG, Planegg/Steinkirchen, Germany), qLV for detection of LV sequences was based on a previous publication by Amendola *et al.*[9] (see Table 1), and the endogene assay qPCBP2 recognised one of the *PCBP2* regions of perfect cross-species sequence identity long enough to allow primer and probe design, and identical amplicon composition (Vector NTI Advance 11.5, Life Technologies) (see Figure 1a). Additional assays qLV-b (FW 5′-TGAAAGCGAAAGGGAAACCA-3′, RV 5′-*TTG*CCGTGCG CGCTTCAG-3′; Probe *HEX*-5′-AGCTCTCTCGACGCAGGACTCGGC-3′-*TAMRA*, based on Miccio *et al.*[12]) and qPCBP2-b (FW 5′-CTGCAT AATCGGGCGTCAAG-3′, RV 5′-GCAGCAGATCCAGTGATGGTAACCT -3′; FAM-5′-CGCCAAAATCAATGAGATCCGTCAGATGTCT-3′-TAMRA, which inadvertently co-amplifies a human pseudogene, *PCBP2P2*, with a single mismatch (underlined) in the probe sequence) were additionally used in simplex reactions and during the optimisation of duplex assays. Attempts to establish a reliable duplex assay for the above range of gDNA and LV amounts included the independent modulation of many PCR parameters, including probe concentration (50, 100, 250 or 375 nM), primer concentration (300, 600 or 900 nM), $MgCl_2$ concentration (3–6 mM), dNTP concentration (200–400 µM), annealing temperature (60, 62 and 64 °C) and the use of alternative qPCR assays qLV-b with qPCBP2-b and qLV with qPCBP2.

All results shown in this study are based on the 2× Multiplex PCR Master Mix (Qiagen), using qPCR assays qLV and qPCBP2 in triplicate 25-µl reactions (1× Master Mix, 250 nM for each probe and 300 nM for each primer), and the 7900HT Fast Real-Time PCR System running SDS v2.4 and the standard PCR programme of 2 min 50 °C, 10 min 95 °C, 40× (15 s 95 °C, 1 min 60 °C) (all from Applied Biosystems). To aid reproduction of the analyses performed here, a simple workflow chart (Supplementary Materials and Methods 1) and Excel (Office 2010, Microsoft, Redmond, WA, USA) spread sheets for the calculation of LV plasmid and gDNA standard curves (Supplementary Materials and Methods 2), for the preparation of reactions for 96-well plates (Supplementary Materials and Methods 3) and for the postrun determination of VCN (Supplementary Materials and Methods 4), respectively, are provided as Supplementary Information. In brief, to avoid bias in the quantification of samples, all threshold cycle (CT) values were determined using automatic calculation of threshold levels in SDS v2.4. All samples were run in triplicate reactions and outliers removed automatically before statistical analyses (all Microsoft Excel), choosing a deviation of 2 s.d. from the triplicate median as a cutoff value. Statistical comparisons (all Microsoft Excel) between expected and observed VCN or gDNA values by duplex assay, and between simplex and duplex results for observed VCN or gDNA values were based on outlier-removed averages of triplicate repeats. For VCN standard curves and quantification of LV insertion sites in cellular clones, relative and absolute VCN values, respectively, were calculated using the ΔΔCT method values with *PCBP2* as the calibrator gene. For analyses of expected versus observed *PCBP2* values, no additional calibrator gene was used and relative *PCBP2* quantities were instead determined by subtracting the smallest *PCBP2* CT value *n* from all the automatically determined CT values x_i of the same run to determine the relative quantity $2^{(xi-n)}$ for each sample. (This can be emulated in Supplementary Materials and Methods 4 by filling in *PCBP2* values for the test gene (column C)] and setting all CT values for the calibrator (column H) to 0).

Amplification efficiencies and the squared Pearson's correlation coefficient r^2 of the curve fit for the log-linear phase of the qPCR amplification were calculated using the program LinRegPCR (v2015.2),[15] by exporting ΔRN values from the SDS v2.4 acquisition software, dismissing outlier-removed reactions and using at least four data points per reaction for fit calculations.

VCNs of LV clones in cell lines were isolated by twofold limiting dilution and selection of clones by visual confirmation of single-cell status after seeding. Absolute copy numbers were determined using standard curves of gDNA dilutions and LV-plasmid dilutions in simplex reactions based on qLV and qPCPB2, and, for MEL cells, additional LV and *PCBP2* qPCR assays. Clones moreover consistently showed the expected LV quantity ratios in relative qPCR comparisons in three independent experiments.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was co-funded by the European Union's Seventh Framework Program for Research, Technological Development and Demonstration under grant agreement number 306201 (THALAMOSS) and by the Republic of Cyprus through the Research Promotion Foundation under grant agreement ΥΓΕΙΑ/ΒΙΟΕ/0311(ΒΕ)/20 and through core funding of the Cyprus Institute of Neurology and Genetics.

REFERENCES

1 Pluta K, Kacprzak MM. Use of HIV as a gene transfer vector. *Acta Biochim Pol* 2009; **56**: 531–595.
2 Cavazza A, Moiani A, Mavilio F. Mechanisms of retroviral integration and mutagenesis. *Hum Gene Ther* 2013; **24**: 119–131.
3 Schuesler T, Reeves L, Kalle C, Grassman E. Copy number determination of genetically-modified hematopoietic stem cells. *Methods Mol Biol* 2009; **506**: 281–298.
4 Cooper AR, Patel S, Senadheera S, Plath K, Kohn DB, Hollis RP. Highly efficient large-scale lentiviral vector concentration by tandem tangential flow filtration. *J Virol Methods* 2011; **177**: 1–9.
5 Romero Z, Urbinati F, Geiger S, Cooper AR, Wherley J, Kaufman ML *et al.* beta-globin gene transfer to human bone marrow for sickle cell disease. *J Clin Invest* 2013.
6 Joglekar AV, Hollis RP, Kuftinec G, Senadheera S, Chan R, Kohn DB. Integrase-defective lentiviral vectors as a delivery platform for targeted modification of adenosine deaminase locus. *Mol Ther* 2013; **21**: 1705–1717.
7 Giannoni F, Hardee CL, Wherley J, Gschweng E, Senadheera S, Kaufman ML *et al.* Allelic exclusion and peripheral reconstitution by TCR transgenic T cells arising from transduced human hematopoietic stem/progenitor cells. *Mol Ther* 2013; **21**: 1044–1054.
8 Karumbayaram S, Lee P, Azghadi SF, Cooper AR, Patterson M, Kohn DB *et al.* From skin biopsy to neurons through a pluripotent intermediate under Good Manufacturing Practice protocols. *Stem Cells Transl Med* 2012; **1**: 36–43.
9 Amendola M, Passerini L, Pucci F, Gentner B, Bacchetta R, Naldini L. Regulated and multiple miRNA and siRNA delivery into primary cells by a lentiviral platform. *Mol Ther* 2009; **17**: 1039–1052.
10 Makeyev AV, Chkheidze AN, Liebhaber SA. A set of highly conserved RNA-binding proteins, alphaCP-1 and alphaCP-2, implicated in mRNA stabilization, are coexpressed from an intronless gene and its intron-containing paralog. *J Biol Chem* 1999; **274**: 24849–24857.
11 Gifford RJ. Viral evolution in deep time: lentiviruses and mammals. *Trends Genet* 2012; **28**: 89–100.
12 Miccio A, Cesari R, Lotti F, Rossi C, Sanvito F, Ponzoni M *et al.* In vivo selection of genetically modified erythroblastic progenitors leads to long-term correction of beta-thalassemia. *Proc Natl Acad Sci USA* 2008; **105**: 10547–10552.
13 Follenzi A, Alles LE, Bakovic S, Geuna M, Naldini L. Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nat Genet* 2000; **25**: 217–222.

14 Roselli EA, Mezzadra R, Fittolli MC, Maruggi G, Biral E, Mavilio F *et al*. Correction of beta-thalassemia major by gene transfer in haematopoietic progenitors of pediatric patients. *EMBO Mol Med* 2010; **2**: 315–328.
15 Ramakers C, Ruijter JM, Deprez RH, Moorman AF. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 2003; **339**: 62–66.

Supplementary Information accompanies this paper on Gene Therapy website (http://www.nature.com/gt)

7 References

1. Hoffman R, Benz EJ, Shattil SJ, et al. Hematology: Basic principles and practice 5th Edition. *Churchill Livingstone Elsevier*. 2009.
2. Becker a J, McCulloch E a, Till JE. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature*. 1963;197(4866):452-454. doi:10.1038/197452a0.
3. Martinez-Agosto JA, Mikkola HKA, Hartenstein V, Banerjee U. The hematopoietic stem cell and its niche: A comparative view. *Genes Dev*. 2007;21(23):3044-3060. doi:10.1101/gad.1602607.
4. Seita J, Weissman IL. Hematopoietic stem cell: Self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med*. 2010;2(6):640-653. doi:10.1002/wsbm.86.
5. Dzierzak E, Philipsen S. Erythropoiesis: development and differentiation. *Cold Spring Harb Perspect Med*. 2013;3(4). doi:10.1101/cshperspect.a011601.
6. McGrath K, Palis J. Ontogeny of Erythropoiesis in the Mammalian Embryo. *Curr Top Dev Biol*. 2008;82(7):1-22. doi:10.1016/S0070-2153(07)00001-4.
7. Palis J. Primitive and definitive erythropoiesis in mammals. *Front Physiol*. 2014;5 JAN. doi:10.3389/fphys.2014.00003.
8. Sankaran VG, Weiss MJ. Anemia: progress in molecular mechanisms and therapies. *Nat Med*. 2015;21(3):221-230. doi:10.1038/nm.3814.
9. Wilkinson AC, Göttgens B. Transcriptional regulation of haematopoietic stem cells. *Adv Exp Med Biol*. 2013;786:187-212. doi:10.1007/978-94-007-6621-1_11.
10. Liu X, Yuan J, Zhang J-W, Zhang X, Wang R. Differential gene expression in human hematopoietic stem cells specified toward erythroid, megakaryocytic, and granulocytic lineage. *J Leukoc Biol*. 2007;82(4):986-1002. doi:10.1189/jlb.0107014.
11. Baron MH. Concise review: Early embryonic erythropoiesis: Not so primitive after all. *Stem Cells*. 2013;31(5):849-856. doi:10.1002/stem.1342.
12. Palis J, Malik J, McGrath KE, Kingsley PD. Primitive erythropoiesis in the mammalian embryo. *Int J Dev Biol*. 2010;54(6-7):1011-1018. doi:10.1387/ijdb.093056jp.
13. McGrath KE, Kingsley PD, Koniski AD, Porter RL, Bushnell TP, Palis J. Enucleation of primitive erythroid cells generates a transient population of “pyrenocytes” in the mammalian fetus. *Blood*. 2008;111(4):2409-2417. doi:10.1182/blood-2007-08-107581.
14. Migliaccio AR, Campisi S, Migliaccio G. Standardization of progenitor cell assay for cord blood banking. *Ann dell'Istituto Super di sanità*. 2001;37(4):595-600.
15. Liu J, Zhang J, Ginzburg Y, et al. Quantitative analysis of murine terminal erythroid differentiation in vivo: novel method to study normal and disordered erythropoiesis. *Blood*. 2013;121(8):e43-9. doi:10.1182/blood-2012-09-456079.
16. Chen K, Liu J, Heck S, Chasis JA, An X, Mohandas N. Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis. *Proc Natl Acad Sci U S A*. 2009;106(41):17413-17418. doi:10.1073/pnas.0909296106.
17. Chasis JA, Mohandas N. Erythroblastic islands: Niches for erythropoiesis. *Blood*. 2008;112(3):470-478. doi:10.1182/blood-2008-03-077883.
18. JM B, JL T, L S. *Hemoglobin Transports Oxygen Efficiently by Binding Oxygen Cooperatively*. W H Freeman; 2002:Section 10.2.
19. Sankaran VG, Orkin SH. The switch from fetal to adult hemoglobin. *Cold Spring Harb Perspect Med*. 2013;3(1):a011643. doi:10.1101/cshperspect.a011643.
20. Boisset J-C, van Cappellen W, Andrieu-Soler C, Galjart N, Dzierzak E, Robin C. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*. 2010;464(7285):116-120. doi:10.1038/nature08764.
21. Giegé R. A historical perspective on protein crystallization from 1840 to the present day. *FEBS J*. 2013;280(24):6456-6497. doi:10.1111/febs.12580.
22. Thein SL. Milestones in the History of Hemoglobin Research (In Memory of Professor Titus H.J. Huisman). *Hemoglobin*. 2011;35(5-6):450-462. doi:10.3109/03630269.2011.613506.
23. Schechter AN. Hemoglobin research and the origins of molecular medicine. *Blood*. 2008;112(10):3927-3938. doi:10.1182/blood-2008-04-078188.
24. Erickson HP. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online*. 2009;11(1):32-51. doi:10.1007/s12575-009-9008-x.
25. Higgs DR, Garrick D, Anguita E, et al. Understanding α-globin gene regulation: Aiming to improve

the management of thalassemia. *Ann N Y Acad Sci*. 2005;1054(1):92-102. doi:10.1196/annals.1345.012.
26. Mettananda S, Gibbons RJ, Higgs DR. α-Globin As a Molecular Target in the Treatment of β-Thalassemia. *Blood*. 2015;125(24):3694-3701. doi:10.1182/blood-2015-03-633594.
27. Vernimmen D. Uncovering Enhancer Functions Using the α-Globin Locus. Rada-Iglesias A, ed. *PLoS Genet*. 2014;10(10):e1004668. doi:10.1371/journal.pgen.1004668.
28. Flint J, Thomas K, Micklem G, et al. The relationship between chromosome structure and function at a human telomeric region. *Nat Genet*. 1997;15(3):252-257. doi:10.1038/ng0397-252.
29. Higgs DR, Engel JD, Stamatoyannopoulos G. Thalassaemia. *Lancet*. 2012;379(9813):373-383. doi:10.1016/S0140-6736(11)60283-3.
30. Gross D. Nuclease Hypersensitive Sites In Chromatin. *Annu Rev Biochem*. 1988;57(1):159-197. doi:10.1146/annurev.biochem.57.1.159.
31. Sankaran VG, Nathan DG. Reversing the hemoglobin switch. *N Engl J Med*. 2010;363(23):2258-2260. doi:10.1056/NEJMcibr1010767.
32. Levings PP, Bungert J. The human β-globin locus control region. A center of attraction. *Eur J Biochem*. 2002;269(6):1589-1599. doi:10.1046/j.1432-1327.2002.02797.x.
33. Bungert J, Tanimoto K, Patel S, Liu Q, Fear M, Engel JD. Hypersensitive site 2 specifies a unique function within the human beta-globin locus control region to stimulate globin gene transcription. *Mol Cell Biol*. 1999;19(4):3062-3072.
34. Noordermeer D, De Laat W. Joining the loops: β-globin gene regulation. *IUBMB Life*. 2008;60(12):824-833. doi:10.1002/iub.129.
35. Cavallesco C, Forget BG, DeRiel JK, Wilson LB, Wilson JT, Weissman SM. Nucleotide sequence of human (G)γ globin messenger RNA. *Gene*. 1980;12(3-4):215-221. doi:10.1016/0378-1119(80)90103-1.
36. Ristaldi MS, Casula S, Porcu S, Marongiu MF, Pirastu M, Cao a. Activation of the delta-globin gene by the beta-globin gene CACCC motif. *Blood Cells Mol Dis*. 1999;25(3-4):193-209. doi:S1079979699902450 [pii].
37. Manchinu MF, Marongiu MF, Poddie D, et al. In vivo activation of the human δ-globin gene: the therapeutic potential in β-thalassemic mice. *Haematologica*. 2014;99(1):76-84. doi:10.3324/haematol.2012.082768.
38. He Z, Russell JE. Expression, purification, and characterization of human hemoglobins Gower-1 (zeta2epsilon2), Gower-2 (alpha2epsilon2), and Portland-2 (zeta2beta2) assembled in complex transgenic-knockout mice. *Blood*. 2001;97(4):1099-1105. doi:10.1182/blood.V97.4.1099.
39. Finotti A, Breda L, Lederer CW, et al. Recent trends in the gene therapy of β-thalassemia. *J Blood Med*. 2015;6:69-85. doi:10.2147/JBM.S46256.
40. Huisman TH. Levels of Hb A2 in heterozygotes and homozygotes for beta-thalassemia mutations: influence of mutations in the CACCC and ATAAA motifs of the beta-globin gene promoter. *Acta Haematol*. 1997;98(4):187-194.
41. Gilman JG. The 12.6 kilobase DNA deletion in Dutch beta zero-thalassaemia. *Br J Haematol*. 1987;67(3):369-372.
42. Bender MA, Ragoczy T, Lee J, et al. The hypersensitive sites of the murine β-globin locus control region act independently to affect nuclear localization and transcriptional elongation. *Blood*. 2012;119(16):3820-3827. doi:10.1182/blood-2011-09-380485.
43. Codrington JF, Li HW, Kutlar F, Gu LH, Ramachandran M, Huisman TH. Observations on the levels of Hb A2 in patients with different beta-thalassemia mutations and a delta chain variant. *Blood*. 1990;76(6):1246-1249.
44. Schokker RC, Went LN, Bok J. A new genetic variant of beta-thalassaemia. *Nature*. 1966;209(5018):44-46.
45. Craig JE, Kelly SJ, Barnetson R, Thein SL. Molecular characterization of a novel 10.3 kb deletion causing beta-thalassaemia with unusually high Hb A2. *Br J Haematol*. 1992;82(4):735-744.
46. Weatherall DJ. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet*. 2001;2(4):245-255. doi:10.1038/35066048.
47. Kohne E. Hemoglobinopathies: clinical manifestations, diagnosis, and treatment. *Dtsch Ärzteblatt Int*. 2011;108(31-32):532-540. doi:10.3238/arztebl.2011.0532.
48. Weatherall D. 2003 William Allan Award address. The Thalassemias: the role of molecular genetics in an evolving global health problem. *Am J Hum Genet*. 2004;74(3):385-392. doi:Doi 10.1086/381402.
49. Breda L, Gambari R, Rivella S. Gene therapy in thalassemia and hemoglobinopathies. *Mediterr J Hematol Infect Dis*. 2009;1(1):e2009008. doi:10.4084/MJHID.2009.008.

50. Piel FB, Patil AP, Howes RE, et al. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat Commun*. 2010;1(8):104. doi:10.1038/ncomms1104.
51. Piel FB, Howes RE, Patil AP, et al. The distribution of haemoglobin C and its prevalence in newborns in Africa. *Sci Rep*. 2013;3:1671. doi:10.1038/srep01671.
52. Higgs DR. The Molecular Basis of α-Thalassemia. *Cold Spring Harb Perspect Med*. 2013;3(1):a011718-a011718. doi:10.1101/cshperspect.a011718.
53. Vichinsky EP. Clinical manifestations of α-thalassemia. *Cold Spring Harb Perspect Med*. 2013;3(5):a011742-a011742. doi:10.1101/cshperspect.a011742.
54. Kountouris P, Lederer CW, Fanis P, Feleki X, Old J, Kleanthous M. IthaGenes: An interactive database for haemoglobin variations and epidemiology. de Brevern AG, ed. *PLoS One*. 2014;9(7):e103020. doi:10.1371/journal.pone.0103020.
55. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. *Bull World Health Organ*. 2008;86(6):480-487. doi:10.2471/BLT.06.036673.
56. Thein SL. The molecular basis of β-thalassemia. *Cold Spring Harb Perspect Med*. 2013;3(5):a011718. doi:10.1101/cshperspect.a011700.
57. Galanello R, Origa R. Beta-thalassemia. *Orphanet J Rare Dis*. 2010;5(1):11. doi:10.1186/1750-1172-5-11.
58. Higgs DR. Ham-Wasserman Lecture: Gene Regulation in Hematopoiesis: New Lessons from Thalassemia. *Hematology*. 2004;2004(1):1-13. doi:10.1182/asheducation-2004.1.1.
59. Yu C, Niakan KK, Matsushita M, Stamatoyannopoulos G, Orkin SH, Raskind WH. X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood*. 2002;100(6):2040-2045. doi:10.1182/blood-2002-02-0387.
60. Peixeiro I, Silva AL, Romão L. Control of human β-globin mRNA stability and its impact on beta-thalassemia phenotype. *Haematologica*. 2011;96(6):905-913. doi:10.3324/haematol.2010.039206.
61. Thein SL. Genetic insights into the clinical diversity of β thalassaemia. *Br J Haematol*. 2004;124(3):264-274. doi:10.1046/j.1365-2141.2003.04769.x.
62. Centis F, Tabellini L, Lucarelli G, et al. The importance of erythroid expansion in determining the extent of apoptosis in erythroid precursors in patients with beta-thalassemia major. *Blood*. 2000;96(10):3624-3629. doi:10.1002/cyto.990150104.
63. Atanasiu V, Manolescu B, Stoian I. Hepcidin - Central regulator of iron metabolism. *Eur J Haematol*. 2007;78(1):1-10. doi:10.1111/j.1600-0609.2006.00772.x.
64. Borgna-Pignatti C, Cappellini MD, De Stefano P, et al. Survival and complications in thalassemia. *Ann N Y Acad Sci*. 2005;1054:40-47. doi:10.1196/annals.1345.006.
65. Cao A, Moi P, Galanello R. Recent advances in β-thalassemias. *Pediatr Rep*. 2011;3(2):e17. doi:10.4081/pr.2011.e17.
66. Dong A, Rivella S, Breda L. Gene therapy for hemoglobinopathies: Progress and challenges. *Transl Res*. 2013;161(4):293-306. doi:10.1016/j.trsl.2012.12.011.
67. Uda M, Galanello R, Sanna S, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of -thalassemia. *Proc Natl Acad Sci*. 2008;105(5):1620-1625. doi:10.1073/pnas.0711566105.
68. Borg J, Papadopoulos P, Georgitsi M, et al. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet*. 2010;42(9):801-805. doi:10.1038/ng.630.
69. Xu J, Sankaran VG, Ni M, et al. Transcriptional silencing of β-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev*. 2010;24(8):783-789. doi:10.1101/gad.1897310.
70. Yu X, Kong Y, Dore LC, et al. An erythroid chaperone that facilitates folding of α-globin subunits for hemoglobin synthesis. *J Clin Invest*. 2007;117(7):1856-1865. doi:10.1172/JCI31664.
71. Kihm AJ, Kong Y, Hong W, et al. An abundant erythroid protein that stabilizes free alpha-haemoglobin. *Nature*. 2002;417(6890):758-763. doi:10.1038/nature00803.
72. Yuan J, Kannan R, Shinar E, Rachmilewitz E a, Low PS. Isolation, characterization, and immunoprecipitation studies of immune complexes from membranes of beta-thalassemic erythrocytes. *Blood*. 1992;79(11):3007-3013.
73. Hocine S, Singer RH, Grünwald D. RNA processing and export. *Cold Spring Harb Perspect Biol*. 2010;2(12):a000752-a000752. doi:10.1101/cshperspect.a000752.
74. Ramanathan A, Robb GB, Chan S-H. mRNA capping: biological functions and applications. *Nucleic Acids Res*. June 2016:gkw551. doi:10.1093/nar/gkw551.

75. Zhang L, Li X, Zhao R. Structural analyses of the pre-mRNA splicing machinery. *Protein Sci*. 2013;22(6):677-692. doi:10.1002/pro.2266.
76. Will CL, Lührmann R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol*. 2011;3(7):1-2. doi:10.1101/cshperspect.a003707.
77. Wahl MC, Will CL, Lührmann R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*. 2009;136(4):701-718. doi:10.1016/j.cell.2009.02.009.
78. De Conti L, Baralle M, Buratti E. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip Rev RNA*. 2013;4(1):49-60. doi:10.1002/wrna.1140.
79. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*. 2009;10(11):741-754. doi:10.1038/nrm2777.
80. Proudfoot N. New perspectives on connecting messenger RNA 3(end formation to transcription. *Curr Opin Cell Biol*. 2004;16(3):272-278. doi:10.1016/j.ceb.2004.03.007.
81. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet*. 2015;17(1):19-32. doi:10.1038/nrg.2015.3.
82. Busslinger M, Moschonas N, Flavell RA. β+ Thalassemia: Aberrant splicing results from a single point mutation in an intron. *Cell*. 1981;27(2 PART 1):289-298. doi:10.1016/0092-8674(81)90412-8.
83. Westaway D, Williamson R. An intron nucleotide sequence variant in a cloned 0+-thassaemia globin gene. *Nucleic Acids Res*. 1981;9(8):1777-1788. doi:10.1093/nar/9.8.1777.
84. Cao A, Galanello R. Beta-thalassemia. *Genet Med*. 2010;12(2):61-76. doi:10.1097/GIM.0b013e3181cd68ed.
85. Jiang Y, Xu X-S, Russell JE. A Nucleolin-Binding 3' Untranslated Region Element Stabilizes -Globin mRNA In Vivo. *Mol Cell Biol*. 2006;26(6):2419-2429. doi:10.1128/MCB.26.6.2419-2429.2006.
86. Hug N, Longman D, Cáceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res*. 2015;44(4):1483-1495. doi:10.1093/nar/gkw010.
87. Khandros E, Thom CS, D'Souza J, Weiss MJ. Integrated protein quality-control pathways regulate free α-globin in murine β-thalassemia. *Blood*. 2012;119(22):5265-5275. doi:10.1182/blood-2011-12-397729.
88. Holbrook J a, Neu-Yilik G, Hentze MW, Kulozik AE. Nonsense-mediated decay approaches the clinic. *Nat Genet*. 2004;36(8):801-808. doi:10.1038/ng1403.
89. Khajavi M, Inoue K, Lupski JR. Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. *Eur J Hum Genet*. 2006;14(10):1074-1081. doi:10.1038/sj.ejhg.5201649.
90. Vadolas J, Nefedov M, Wardan H, et al. Humanized β-thalassemia mouse model containing the common IVSI-110 splicing mutation. *J Biol Chem*. 2006;281(11):7399-7405. doi:10.1074/jbc.M512931200.
91. Dominski Z, Kole R. Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. *Pnas*. 1996;58(June):359-381. doi:10.1073/pnas.90.18.8673.
92. Reed R, Maniatis T. Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell*. 1985;41(1):95-105. doi:10.1016/0092-8674(85)90064-9.
93. Baysal E, Indrak K, Bozkurt G, et al. The β-thalassaemia mutations in the population of Cyprus. *Br J Haematol*. 1992;81(4):607-609.
94. El-Beshlawy A, Mostafa A, Youssry I, et al. Correction of Aberrant Pre-mRNA Splicing by Antisense Oligonucleotides in b -Thalassemia Egyptian Patients. *J Pediatr Hematol Oncol*. 2008;30(4):281-284. doi:10.1097/MPH.0b013e3181639afe.
95. Vacek MM, Ma H, Gemignani F, Lacerra G, Kafri T, Kole R. High-level expression of hemoglobin A in human thalassemic erythroid progenitor cells following lentiviral vector delivery of an antisense snRNA. *Blood*. 2003;101(1):104-111. doi:10.1182/blood-2002-06-1869.
96. Vacek M, Sazani P, Kole R. Antisense-mediated redirection of mRNA splicing. *Cell Mol Life Sci*. 2003;60(5):825-833. doi:10.1007/s00018-003-3042-7.
97. Sadelain M, Boulad F, Galanello R, et al. Therapeutic options for patients with severe beta-thalassemia: the need for globin gene therapy. *Hum Gene Ther*. 2007;18(1):1-9. doi:10.1089/hum.2006.151.
98. Gluckman E. Ten years of cord blood transplantation: From bench to bedside. *Br J Haematol*. 2009;147(2):192-199. doi:10.1111/j.1365-2141.2009.07780.x.
99. Pinto FO, Roberts I. Cord blood stem cell transplantation for haemoglobinopathies. *Br J Haematol*. 2008;141(3):309-324. doi:10.1111/j.1365-2141.2008.07016.x.
100. Norkin M, Lazarus HM, Wingard JR. Umbilical cord blood graft enhancement strategies: has the time come to move these into the clinic? *Bone Marrow Transplant*. 2013;48(7):884-889. doi:10.1038/bmt.2012.163.

101. Locatelli F, Merli P, Strocchio L. Transplantation for thalassemia major. *Curr Opin Hematol*. August 2016:1. doi:10.1097/MOH.0000000000000280.
102. Lan WS, Phak NH, Ying TP, et al. Management of Thalasaemia. *Heal Technol Assesment Unit, Med Dev Div Minist Heal Malaysia*. 2003;3(MOH/PAK/77.03 (TR)):137.
103. Mabaera R, West RJ, Conine SJ, et al. A cell stress signaling model of fetal hemoglobin induction: what doesn't kill red blood cells may make them stronger. *Exp Hematol*. 2008;36(9):1057-1072. doi:10.1016/j.exphem.2008.06.014.
104. Sauntharajah Y, Hillery CA, Lavelle D, et al. Effects of 5-aza-2 '-deoxycytidine on fetal hemoglobin levels, red cell adhesion, and hematopoietic differentiation in patients with sickle cell disease. *Blood*. 2003;102(12):3865-3870. doi:DOI 10.1182/blood-2003-05-1738.
105. Testa U. Fetal hemoglobin chemical inducers for treatment of hemoglobinopathies. *Ann Hematol*. 2009;88(6):505-528. doi:10.1007/s00277-008-0637-y.
106. Pace BS, Zein S. Understanding mechanisms of γ-globin gene regulation to develop strategies for pharmacological fetal hemoglobin induction. *Dev Dyn*. 2006;235(7):1727-1737. doi:10.1002/dvdy.20802.
107. Mort M, Ivanov D, Cooper DN, Chuzhanova NA. A meta-analysis of nonsense mutations causing human genetic disease. *Hum Mutat*. 2008;29(8):1037-1047. doi:10.1002/humu.20763.
108. Altamura E, Borgatti M, Finotti A, et al. Chemical-Induced Read-Through at Premature Termination Codons Determined by a Rapid Dual-Fluorescence System Based on S. cerevisiae. *PLoS One*. 2016;11(4):e0154260. doi:10.1371/journal.pone.0154260.
109. Lee HLR, Dougherty JP. *Pharmaceutical Therapies to Recode Nonsense Mutations in Inherited Diseases*. Vol 136.; 2012:227-266. doi:10.1016/j.pharmthera.2012.07.007.
110. Bidou L, Allamand V, Rousset JP, Namy O. Sense from nonsense: Therapies for premature stop codon diseases. *Trends Mol Med*. 2012;18(11):679-688. doi:10.1016/j.molmed.2012.09.008.
111. Keeling KM, Xue X, Gunn G, Bedwell DM. Therapeutics Based on Stop Codon Readthrough. *Annu Rev Genomics Hum Genet*. 2014;(April):1-24. doi:10.1146/annurev-genom-091212-153527.
112. Salvatori F, Breveglieri G, Zuccato C, et al. Production of β-globin and adult hemoglobin following G418 treatment of erythroid precursor cells from homozygous β039 thalassemia patients. *Am J Hematol*. 2009;84(11):720-728. doi:10.1002/ajh.21539.
113. Du L, Damoiseaux R, Nahas S, et al. Nonaminoglycoside compounds induce readthrough of nonsense mutations. *J Exp Med*. 2009;206(10):2285-2297. doi:10.1084/jem.20081940.
114. Breda L, Casu C, Gardenghi S, et al. Therapeutic hemoglobin levels after gene transfer in β-thalassemia mice and in hematopoietic cells of β-thalassemia and sickle cells disease patients. *PLoS One*. 2012;7(3):e32345. doi:10.1371/journal.pone.0032345.
115. Roselli EA, Mezzadra R, Frittoli MC, et al. Correction of β-thalassemia major by gene transfer in haematopoietic progenitors of pediatric patients. *EMBO Mol Med*. 2010;2(8):315-328. doi:10.1002/emmm.201000083.
116. Kotterman MA, Chalberg TW, Schaffer D V. Viral Vectors for Gene Therapy: Translational and Clinical Outlook. *Annu Rev Biomed Eng*. 2015;17(1):63-89. doi:10.1146/annurev-bioeng-071813-104938.
117. Drakopoulou E, Papanikolaou E, Georgomanoli M, Anagnou NP. Towards more successful gene therapy clinical trials for β-thalassemia. *Curr Mol Med*. 2013;13(8):1314-1330. doi:10.2174/15665240113139990064.
118. Chandrakasan S, Malik P. Gene therapy for hemoglobinopathies: The state of the field and the future. *Hematol Oncol Clin North Am*. 2014;28(2):199-216. doi:10.1016/j.hoc.2013.12.003.
119. Kutner RH, Zhang XY, Reiser J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat Protoc*. 2009;4(4):495-505. doi:10.1038/nprot.2009.22.
120. Chen ST, Iida a, Guo L, Friedmann T, Yee JK. Generation of packaging cell lines for pseudotyped retroviral vectors of the G protein of vesicular stomatitis virus by using a modified tetracycline inducible system. *Proc Natl Acad Sci U S A*. 1996;93(September):10057-10062. doi:10.1073/pnas.93.19.10057.
121. Imren S, Payen E, Westerman KA, et al. Permanent and panerythroid correction of murine beta thalassemia by multiple lentiviral integration in hematopoietic stem cells. *Proc Natl Acad Sci U S A*. 2002;99(22):14380-14385. doi:10.1073/pnas.212507099.
122. Moi P, Sadelain M. Towards the genetic treatment of β-thalassemia: New disease models, new vectors, new cells. *Haematologica*. 2008;93(3):325-330. doi:10.3324/haematol.12732.
123. Persons DA, Hargrove PW, Allay ER, Hanawa H, Nienhuis AW. The degree of phenotypic correction of murine β-thalassemia intermedia following lentiviral-mediated transfer of a human γ-globin gene is influenced by chromosomal position effects and vector copy number. *Blood*.

2003;101(6):2175-2183. doi:10.1182/blood-2002-07-2211.
124. Pestina TI, Hargrove PW, Jay D, Gray JT, Boyd KM, Persons DA. Correction of murine sickle cell disease using gamma-globin lentiviral vectors to mediate high-level expression of fetal hemoglobin. *Mol Ther*. 2009;17(2):245-252. doi:mt2008259 [pii]\r10.1038/mt.2008.259.
125. Rivella S, May C, Chadburn A, Rivière I, Sadelain M. A novel murine model of Cooley anemia and its rescue by lentiviral-mediated human β-globin gene transfer. *Blood*. 2003;101(8):2932-2939. doi:10.1182/blood-2002-10-3305.
126. May C, Rivella S, Callegari J, et al. Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. *Nature*. 2000;406(6791):82-86. doi:10.1038/35017565.
127. May C. Successful treatment of murine beta -thalassemia intermedia by transfer of the human beta -globin gene. *Blood*. 2002;99(6):1902-1908. doi:10.1182/blood.V99.6.1902.
128. Persons DA, Allay ER, Sabatino DE, Kelly P, Bodine DM, Nienhuis AW. Functional requirements for phenotypic correction of murine β-thalassemia: Implications for human gene therapy. *Blood*. 2001;97(10):3275-3282. doi:10.1182/blood.V97.10.3275.
129. Arumugam P, Malik P. Genetic Therapy for Beta-Thalassemia: From the Bench to the Bedside. *Hematology*. 2010;2010(1):445-450. doi:10.1182/asheducation-2010.1.445.
130. Pawliuk R, Westerman KA, Fabry ME, et al. Correction of Sickle Cell Disease in Transgenic Mouse Models by Gene Therapy. *Science (80-)*. 2001;294(5550):2368-2371. doi:10.1126/science.1065806.
131. Negre O, Eggimann A-V, Beuzard Y, et al. Gene Therapy of the β-Hemoglobinopathies by Lentiviral Transfer of the β(A(T87Q))-Globin Gene. *Hum Gene Ther*. 2016;27(2):148-165. doi:10.1089/hum.2016.007.
132. Miccio A, Cesari R, Lotti F, et al. In vivo selection of genetically modified erythroblastic progenitors leads to long-term correction of beta-thalassemia. *Proc Natl Acad Sci U S A*. 2008;105(30):10547-10552. doi:10.1073/pnas.0711666105.
133. Angelucci E, Matthes-Martin S, Baronciani D, et al. Hematopoietic stem cell transplantation in thalassemia major and sickle cell disease: Indications and management recommendations from an international expert panel. *Haematologica*. 2014;99(5):811-820. doi:10.3324/haematol.2013.099747.
134. Payen E, Leboulch P. Advances in stem cell transplantation and gene therapy in the β-hemoglobinopathies. *Hematology Am Soc Hematol Educ Program*. 2012;2012(1):276-283. doi:10.1182/asheducation-2012.1.276.
135. Miccio A, Poletti V, Tiboni F, et al. The GATA1-HS2 enhancer allows persistent and position-independent expression of a β-globin transgene. *PLoS One*. 2011;6(12):e27955. doi:10.1371/journal.pone.0027955.
136. Bank A, Dorazio R, Leboulch P. A Phase I/II Clinical Trial of β-Globin Gene Therapy for β-Thalassemia. *Ann N Y Acad Sci*. 2005;1054(1):308-316. doi:10.1196/annals.1345.007.
137. Cavazzana-Calvo M, Payen E, Negre O, et al. Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. *Nature*. 2010;467(7313):318-322. doi:10.1038/nature09328.
138. Kaiser J. Gene therapy. Beta-thalassemia treatment succeeds, with a caveat. *Science*. 2009;326(5959):1468-1469. doi:10.1126/science.326.5959.1468-b.
139. Cavazzana M, Ribeil J-A, Payen E, et al. Outcomes of gene therapy for beta-thalassaemia major via transplantation of autologous hematopoietic stem cells transduced ex vivo with lentiviral beta globin vector. *Hum Gene Ther*. 2014.
140. Morishita A, Zaidi MR, Mitoro A, et al. HMGA2 is a driver of tumor metastasis. *Cancer Res*. 2013;73(14):4289-4299. doi:10.1158/0008-5472.CAN-12-3848.
141. Ronen K, Negre O, Roth S, et al. Distribution of lentiviral vector integration sites in mice following therapeutic gene transfer to treat β-thalassemia. *Mol Ther*. 2011;19(7):1273-1286. doi:10.1038/mt.2011.20.
142. Desprat R, Bouhassira EE. Gene specificity of suppression of transgene-mediated insertional transcriptional activation by the chicken HS4 insulator. *PLoS One*. 2009;4(6):e5956. doi:10.1371/journal.pone.0005956.
143. Negre O, Bartholomae C, Beuzard Y, et al. Preclinical evaluation of efficacy and safety of an improved lentiviral vector for the treatment of β-thalassemia and sickle cell disease. *Curr Gene Ther*. 2015;15(1):64-81. doi:10.1016/j.ijpharm.2013.11.041.
144. Cavazzana M, Ribeil J, Payen E, Suarez F, Beuzard Y. Outcomes of Gene Therapy for Severe Sickle Disease and Beta-Thalassemia Major Via Transplantation of Autologous Hematopoietic Stem

Cells Transduced Ex Vivo with a Lentiviral Beta AT87Q-Globin Vector. *Blood (ASH Annu Meet Abstr*. 2015;126(23):1-6.
145. Cavazzana M, Ribeil J-A, Payen E, et al. Study Hgb-205: Outcomes of Gene Therapy for Hemoglobinopathies Via Transplantation of Autologous Hematopoietic Stem Cells Transduced Ex Vivo with a Lentiviral βA-T87Q-Globin Vector (LentiGlobin® BB305 Drug Product). *Blood*. 2014;124(21):4797-4797.
146. Cavazzana M, Ribeil J-A, Payen E, et al. Outcomes of Gene Therapy for Severe Sickle Disease and Beta-Thalassemia Major Via Transplantation of Autologous Hematopoietic Stem Cells Transduced Ex Vivo with a Lentiviral Beta AT87Q-Globin Vector. *Blood*. 2015;126(23):202-202.
147. Boulad F, Wang X, Qu J, et al. Safe mobilization of CD34+ cells in adults with β-thalassemia and validation of effective globin gene transfer for clinical investigation. *Blood*. 2014;123(10):1483-1486. doi:10.1182/blood-2013-06-507178.
148. Zonari E, Meo O, Scaramuzza S, et al. Improved Ex Vivo Gene Therapy Using Highly Purified Hematopoietic Stem and Progenitor Cells. *ASGCT, 19th Annu Meet Washinght DC*. 2016.
149. Ramachandran PV, Ignacimuthu S. RNA interference - A silent but an efficient therapeutic tool. *Appl Biochem Biotechnol*. 2013;169(6):1774-1789. doi:10.1007/s12010-013-0098-1.
150. Liu YP, Berkhout B. MiRNA cassettes in viral vectors: Problems and solutions. *Biochim Biophys Acta - Gene Regul Mech*. 2011;1809(11-12):732-745. doi:10.1016/j.bbagrm.2011.05.014.
151. Boudreau RL, Martins I, Davidson BL. Artificial microRNAs as siRNA shuttles: improved safety as compared to shRNAs in vitro and in vivo. *Mol Ther*. 2009;17(1):169-175. doi:10.1038/mt.2008.231.
152. McBride JL, Boudreau RL, Harper SQ, et al. Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: implications for the therapeutic development of RNAi. *Proc Natl Acad Sci U S A*. 2008;105(15):5868-5873. doi:10.1073/pnas.0801775105.
153. Bauer M, Kinkl N, Meixner a, et al. Prevention of interferon-stimulated gene expression using microRNA-designed hairpins. *Gene Ther*. 2009;16(1):142-147. doi:10.1038/gt.2008.123.
154. Voon HPJ, Wardan H, Vadolas J. Co-inheritance of α- and β-thalassaemia in mice ameliorates thalassaemic phenotype. *Blood Cells, Mol Dis*. 2007;39(2):184-188. doi:10.1016/j.bcmd.2007.01.006.
155. Voon HPJ, Wardan H, Vadolas J. siRNA-mediated reduction of alpha-globin results in phenotypic improvements in beta-thalassemic cells. *Haematologica*. 2008;93(8):1238-1242. doi:10.3324/haematol.12555.
156. Xie SY, Ren ZR, Zhang JZ, et al. Restoration of the balanced α/β-globin gene expression in β654-thalassemia mice using combined RNAi and antisense RNA approach. *Hum Mol Genet*. 2007;16(21):2616-2625. doi:10.1093/hmg/ddm218.
157. Xie SY, Li W, Ren ZR, Huang SZ, Zeng F, Zeng YT. Correction of β654-thalassaemia mice using direct intravenous injection of siRNA and antisense RNA vectors. *Int J Hematol*. 2011;93(3):301-310. doi:10.1007/s12185-010-0727-1.
158. Samakoglu S, Lisowski L, Budak-Alpdogan T, et al. A genetic strategy to treat sickle cell anemia by coregulating globin transgene expression and RNA interference. *Nat Biotechnol*. 2006;24(1):89-94. doi:10.1038/nbt1176.
159. Xu J, Bauer DE, Orkin SH. Targeting Regulators of Hemoglobin F. *Hematol*. 2011.
160. Brendel C, Guda S, Renella R, et al. Lineage-specific BCL11A knockdown circumvents toxicities and reverses sickle phenotype. *J Clin Invest*. 2016;126(10):1-11. doi:10.1172/JCI87885.Curative.
161. Kumar SR, Markusic DM, Biswas M, High KA, Herzog RW. Clinical development of gene therapy: results and lessons from recent successes. *Mol Ther Methods Clin Dev*. 2016;3:16034. doi:10.1038/mtm.2016.34.
162. Ylä-Herttuala S. ADA-SCID Gene Therapy Endorsed by European Medicines Agency for Marketing Authorization. *Mol Ther*. 2016;24(6):1013-1014. doi:10.1038/mt.2016.98.
163. Cox DBT, Platt RJ, Zhang F. Therapeutic genome editing: prospects and challenges. *Nat Med*. 2015;21(2):121-131. doi:10.1038/nm.3793.
164. Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*. 1985;317(6034):230-234. doi:10.1038/317230a0.
165. Sedivy JM, Dutriaux A. Gene targeting and somatic cell genetics--a rebirth or a coming of age? *Trends Genet*. 1999;15(3):88-90. doi:S0168-9525(98)01689-8 [pii].
166. Porteus MH. Towards a new era in medicine: therapeutic genome editing. *Genome Biol*. 2015;16(1):286-298. doi:10.1186/s13059-015-0859-y.
167. Jasin M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet*.

1996;12(6):224-228. doi:10.1016/0168-9525(96)10019-6.
168. Davis AJ, Chen DJ. DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res*. 2013;2(3):130-143. doi:10.3978/j.issn.2218-676X.2013.04.02.
169. Lin S, Staahl B, Alla RK, Doudna J a. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*. 2014;3:1-13. doi:10.7554/eLife.04766.
170. Heyer W-D, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet*. 2010;44:113-139. doi:10.1146/annurev-genet-051710-150955.
171. Papapetrou EP, Lee G, Malani N, et al. Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. *Nat Biotechnol*. 2011;29(1):73-78. doi:10.1038/nbt.1717.
172. Porteus M. Genome Editing: A New Approach to Human Therapeutics. *Annu Rev Pharmacol Toxicol*. 2016;56(1):163-190. doi:10.1146/annurev-pharmtox-010814-124454.
173. Dueva R, Iliakis G. Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer. *Transl Cancer Res*. 2013;2(3):163-177. doi:10.3978/j.issn.2218-676X.2013.05.02.
174. Ousterout DG, Kabadi AM, Thakore PI, Majoros WH, Reddy TE, Gersbach CA. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat Commun*. 2015;6:6244. doi:10.1038/ncomms7244.
175. Ousterout DG, Kabadi AM, Thakore PI, et al. Correction of dystrophin expression in cells from Duchenne muscular dystrophy patients through genomic excision of exon 51 by zinc finger nucleases. *Mol Ther*. 2015;23(3):523-532. doi:10.1038/mt.2014.234.
176. Ousterout DG, Perez-Pinera P, Thakore PI, et al. Reading frame correction by targeted genome editing restores dystrophin expression in cells from duchenne muscular dystrophy patients. *Mol Ther*. 2013;21(9):1718-1726. doi:10.1038/mt.2013.111.
177. Tebas P, Stein D, Ww T, et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med*. 2014;370(10):901-910. doi:10.1056/NEJMoa1300662.
178. Maeder ML, Gersbach CA. Genome-editing Technologies for Gene and Cell Therapy. *Mol Ther*. 2016;24(3):430-446. doi:10.1038/mt.2016.10.
179. Gaj T, Gersbach CA, Barbas CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. 2013;31(7):397-405. doi:10.1016/j.tibtech.2013.04.004.
180. Glazkova D V, Shipulin GA. Tale nucleases as a new tool for genome editing. *Mol Biol*. 2014;48(3):355-370. doi:10.1134/S0026893314030054.
181. Kim Y, Kweon J, Kim J-S. TALENs and ZFNs are associated with different mutation signatures. *Nat Methods*. 2013;10(3):185. doi:10.1038/nmeth.2364.
182. Streubel J, Blücher C, Landgraf A, Boch J. TAL effector RVD specificities and efficiencies. *Nat Biotechnol*. 2012;30(7):593-595. doi:10.1038/nbt.2304.
183. Boch J, Scholze H, Schornack S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*. 2009;326(5959):1509-1512. doi:10.1126/science.1178811.
184. Miller JC, Zhang L, Xia DF, et al. Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat Methods*. 2015;12(5):465-471. doi:10.1038/nmeth.3330.
185. Mussolino C, Cathomen T. TALE nucleases: Tailored genome engineering made easy. *Curr Opin Biotechnol*. 2012;23(5):644-650. doi:10.1016/j.copbio.2012.01.013.
186. Miller JC, Tan S, Qiao G, et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*. 2011;29(2):143-148. doi:10.1038/nbt.1755.
187. Morbitzer R, Elsaesser J, Hausner J, Lahaye T. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res*. 2011;39(13):5790-5799. doi:10.1093/nar/gkr151.
188. Cermak T, Doyle EL, Christian M, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res*. 2011;39(12):e82. doi:10.1093/nar/gkr218.
189. Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*. 2012;30(5):460-465. doi:10.1038/nbt.2170.
190. Hendel A, Fine EJ, Bao G, Porteus MH. Quantifying on- and off-target genome editing. *Trends Biotechnol*. 2015;33(2):132-140. doi:10.1016/j.tibtech.2014.12.001.
191. Holkers M, Maggio I, Liu J, et al. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res*. 2013;41(5):e63. doi:10.1093/nar/gks1446.
192. Yang L, Guell M, Byrne S, et al. Optimization of scarless human stem cell genome editing. *Nucleic Acids Res*. 2013;41(19):9049-9061. doi:10.1093/nar/gkt555.

193. Charpentier E. CRISPR-Cas9: how research on a bacterial RNA-guided mechanism opened new perspectives in biotechnology and biomedicine. *EMBO Mol Med*. 2015;7(4):363-365. doi:10.15252/emmm.201504847.
194. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014;157(6):1262-1278. doi:10.1016/j.cell.2014.05.010.
195. Cong L, Ran FA, Cox D, et al. Multiplex Genome Engineering Using CRISPR/Cas System. *Science (80-)*. 2013;339(February):819-824. doi:10.1126/science.1231143 RNA-Guided.
196. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-)*. 2012;337(August):816-822. doi:10.1126/science.1225829.
197. Mali P, Yang L, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. *Science (80-)*. 2013;339(6121):823-826. doi:10.1126/science.1232033.
198. Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31(9):827-832. doi:10.1038/nbt.2647.
199. Jiang W, Bikard D, Cox D, Zhang F, Marraffini L a. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31(3):233-239. doi:10.1038/nbt.2508.
200. Zetsche B, Gootenberg JS, Abudayyeh OO, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. 2015;163(3):759-771. doi:10.1016/j.cell.2015.09.038.
201. Ran FA, Hsu PD, Lin CY, et al. Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*. 2013;154(6):1380-1389. doi:10.1016/j.cell.2013.08.021.
202. Tsai SQ, Wyvekens N, Khayter C, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*. 2014;32(6):569-576. doi:10.1038/nbt.2908.
203. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*. 2014;32(3):279-284. doi:10.1038/nbt.2808.
204. Guo J, Gaj T, Barbas CF. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J Mol Biol*. 2010;400(1):96-107. doi:10.1016/j.jmb.2010.04.060.
205. Miller JC, Holmes MC, Wang J, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*. 2007;25(7):778-785. doi:10.1038/nbt1319.
206. Canver MC, Smith EC, Sher F, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. 2015;527(7577):192-197. doi:10.1038/nature15521.
207. Vierstra J, Reik A, Chang K-H, et al. Functional footprinting of regulatory DNA. *Nat Methods*. 2015;12(10):927-930. doi:10.1038/nmeth.3554.
208. Canver MC, Orkin SH. Customizing the genome as therapy for the β-hemoglobinopathies. *Blood*. 2016;127(21):blood-2016-01-678128. doi:10.1182/blood-2016-01-678128.
209. Andreani M, Nesci S, Lucarelli G, et al. Long-term survival of ex-thalassemic patients with persistent mixed chimerism after bone marrow transplantation. *Bone Marrow Transplant*. 2000;25(4):401-404. doi:10.1038/sj.bmt.1702151.
210. Voit RA, Hendel A, Pruett-Miller SM, Porteus MH. Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic Acids Res*. 2014;42(2):1365-1378. doi:10.1093/nar/gkt947.
211. Ma N, Liao B, Zhang H, et al. Transcription activator-like effector nuclease (TALEN)-mediated Gene correction in integration-free β-Thalassemia induced pluripotent stem cells. *J Biol Chem*. 2013;288(48):34671-34679. doi:10.1074/jbc.M113.496174.
212. Ramalingam S, Annaluru N, Kandavelou K, Chandrasegaran S. TALEN-mediated generation and genetic correction of disease-specific human induced pluripotent stem cells. *Curr Gene Ther*. 2014;14(6):461-472. doi:10.2174/1566523214666140918101725.
213. Huang X, Wang Y, Yan W, et al. Production of gene-corrected adult beta globin protein in human erythrocytes differentiated from patient ipscs after genome editing of the sickle point mutation. *Stem Cells*. 2015;33(5):1470-1479. doi:10.1002/stem.1969.
214. Xie F, Ye L, Chang JC, et al. Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res*. 2014;24(9):1526-1533. doi:10.1101/gr.173427.114.
215. Sun N, Zhao H. Seamless correction of the sickle cell disease mutation of the HBB gene in human induced pluripotent stem cells using TALENs. *Biotechnol Bioeng*. 2014;111(5):1048-1053. doi:10.1002/bit.25018.
216. Xu P, Tong Y, Liu X, et al. Both TALENs and CRISPR/Cas9 directly target the HBB IVS2-654 (C > T) mutation in β-thalassemia-derived iPSCs. *Sci Rep*. 2015;5(April):12065. doi:10.1038/srep12065.
217. Hoban MD, Cost GJ, Mendel MC, et al. Correction of the sickle-cell disease mutation in human hematopoietic stem/progenitor cells. *Blood*. 2015;125(17):2597-2604. doi:10.1182/blood-2014-

12-615948.
218. Genovese P, Schiroli G, Escobar G, et al. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature*. 2014;510(7504):235-240. doi:10.1038/nature13420.
219. Hoban MD, Lumaquin D, Kuo CY, et al. CRISPR/Cas9-Mediated Correction of the Sickle Mutation in Human CD34+ cells. *Mol Ther*. 2016. doi:10.1038/mt.2016.148.
220. Traxler EA, Yao Y, Wang Y-D, et al. A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat Med*. 2016;22(9):987-990. doi:10.1038/nm.4170.
221. Gallagher PG, Steiner LA, Liem RI, et al. Mutation of a barrier insulator in the human ankyrin-1 gene is associated with hereditary spherocytosis. *J Clin Invest*. 2010;120(12):4453-4465. doi:10.1172/JCI42240.
222. Breda L, Casu C, Casula L, et al. Following Beta-Globin Gene Transfer, the Production of Hemoglobin Depends Upon the Beta-Thalassemia Genotype. *Blood (ASH Annu Meet Abstr*. 2009;51st ASH A(114: Abstract 978):978-978.
223. Breda L, Casu C, Kleinert DA, et al. Following β-globin Gene Transfer, the Production of Hemoglobin Depends upon the β-Thalassemia Genotype: The "Traffic Jam Hypothesis." *12th Annu Meet Am Soc Gene Ther San Diego, Calif*. 2009.
224. Schoenberg DR, Maquat LE. Regulation of cytoplasmic mRNA decay. *Nat Rev Genet*. 2012;13(4):246-259. doi:10.1038/nrg3160.
225. Breda L, Kleinert DA, Casu C, et al. A preclinical approach for gene therapy of β-thalassemia. *Ann N Y Acad Sci*. 2010;1202(1):134-140. doi:10.1111/j.1749-6632.2010.05594.x.
226. Sambrook JR. *DW. 2001 Molecular Cloning: A Laboratory Manual*. 3rd ed. Cold Spring Harbor, N.Y., U.S.A.: Cold Spring Harbor Laboratory Press; 2001.
227. Mussolino C, Alzubi J, Fine EJ, et al. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Res*. 2014;42(10):6762-6773. doi:10.1093/nar/gku305.
228. Mussolino C, Morbitzer R, Lütge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*. 2011;39(21):9283-9293. doi:10.1093/nar/gkr597.
229. Martin P, Papayannopoulou T. HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. *Science*. 1982;216(4551):1233-1235. doi:10.1126/science.6177045.
230. Levenson R, Housman D. Developmental program of murine erythroleukemia cells. Effect of the inhibition of protein synthesis. *J Cell Biol*. 1979;82(3):715-725.
231. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol*. 1977;36(1):59-72. doi:10.1099/0022-1317-36-1-59.
232. Dull T, Zufferey R, Kelly M, et al. A third-generation lentivirus vector with a conditional packaging system. *J Virol*. 1998;72(11):8463-8471. doi:98440501.
233. Promega Corporation. pAdVAntage™ Vector Technical Bulletin. https://worldwide.promega.com/resources/protocols/technical-bulletins/0/padvantage-vector-protocol/. Published 2016.
234. TaqMan ® Universal PCR Master Mix Protocol. 2002.
235. Christodoulou I, Patsali P, Stephanou C, Antoniou M, Kleanthous M, Lederer CW. Measurement of lentiviral vector titre and copy number by cross-species duplex quantitative PCR. *Gene Ther*. 2016;23(1):113-118. doi:10.1038/gt.2015.60.
236. Brinkman EK, Chen T, Amendola M, Van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*. 2014;42(22):e168. doi:10.1093/nar/gku936.
237. Fine EJ, Cradick TJ, Zhao CL, Lin Y, Bao G. An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Res*. 2014;42(6):e42. doi:10.1093/nar/gkt1326.
238. Danckwardt S, Neu-Yilik G, Thermann R, Frede U, Hentze MW, Kulozik AE. Abnormally spliced β-globin mRNAs: A single point mutation generates transcripts sensitive and insensitive to nonsense-mediated mRNA decay. *Blood*. 2002;99(5):1811-1816. doi:10.1182/blood.V99.5.1811.
239. Fukumaki Y, Ghosh PK, Benz EJ, et al. Abnormally spliced messenger RNA in erythroid cells from patients with β+ thalassemia and monkey cells expressing a cloned β+-thalassemic gene. *Cell*. 1982;28(3):585-593. doi:10.1016/0092-8674(82)90213-6.
240. Antoniou M. Induction of Erythroid-Specific Expression in Murine Erythroleukemia (MEL) Cell Lines. *Methods Mol Biol*. 1991;7(10):421-434. doi:10.1385/0-89603-178-0:421.

241. Thein SL. The molecular basis of β-thalassemia. *Cold Spring Harb Perspect Med*. 2013;3(5). doi:10.1101/cshperspect.a011700.
242. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. *Genes Dev*. 2009;23(11):1247-1269. doi:10.1101/gad.1792809.
243. Braglia P, Percudani R, Dieci G. Sequence context effects on oligo(dT) termination signal recognition by Saccharomyces cerevisiae RNA polymerase III. *J Biol Chem*. 2005;280(20):19551-19562. doi:10.1074/jbc.M412238200.
244. Ui-Tei K, Nishi K, Takahashi T, Nagasawa T. Thermodynamic control of small RNA-mediated gene silencing. *Front Genet*. 2012;3(JUN):101. doi:10.3389/fgene.2012.00101.
245. Hutson TH, Foster E, Dawes JM, Hindges R, Yáñez-Muñoz RJ, Moon LDF. Lentiviral vectors encoding short hairpin RNAs efficiently transduce and knockdown LINGO-1 but induce an interferon response and cytotoxicity in central nervous system neurones. *J Gene Med*. 2012;14(5):299-315. doi:10.1002/jgm.2626.
246. Moiani A, Paleari Y, Sartori D, et al. Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *J Clin Invest*. 2012;122(5):1653-1666. doi:10.1172/JCI61852DS1.
247. Ge Y, Porse BT. The functional consequences of intron retention: Alternative splicing coupled to NMD as a regulator of gene expression. *BioEssays*. 2014;36(3):236-243. doi:10.1002/bies.201300156.
248. Cradick TJ, Fine EJ, Antico CJ, Bao G. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res*. 2013;41(20):9584-9592. doi:10.1093/nar/gkt714.
249. Luo Y, Zhu D, Zhang Z, Chen Y, Sun X. Integrative analysis of CRISPR/Cas9 target sites in the human HBB gene. *Biomed Res Int*. 2015;2015:1-9. doi:10.1155/2015/514709.
250. Certo MT, Gwiazda KS, Kuhar R, et al. Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat Methods*. 2012;9(10):973-975. doi:10.1038/nmeth.2177.
251. Kim H, Kim JS. A guide to genome engineering with programmable nucleases. *Nat Rev Genet*. 2014;15(5):321-334. doi:10.1038/nrg3686.
252. Chari R, Mali P, Moosburner M, Church GM. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods*. 2015;12(July):1-7. doi:10.1038/nmeth.3473.
253. Lee Y, Rio DC, Biology S, Biology C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *AnnuRevBiochem*. 2015;(258):291-323. doi:10.1146/annurev-biochem-060614-034316.Mechanisms.
254. Cao a, Moi P. Regulation of the globin genes. *Pediatr Res*. 2002;51(4):415-421. doi:10.1203/00006450-200204000-00003.
255. Lodish HF, Jacobsen M. Regulation of hemoglobin synthesis. Equal rates of translation and termination of α- and β-globin chains. *J Biol Chem*. 1972;247(11):3622-3629.
256. Sadelain M, Rivière I, Wang X, et al. Strategy for a multicenter phase i clinical trial to evaluate globin gene transfer in β-thalassemia. *Ann N Y Acad Sci*. 2010;1202:52-58. doi:10.1111/j.1749-6632.2010.05597.x.
257. Zhang SB, He QY, Zhao H, Gui CY, Jiang C, Qian RL. Function of GATA transcription factors in hydroxyurea-induced HEL cells. *Cell Res*. 2001;11(4):301-310. doi:10.1038/sj.cr.7290100.
258. Wilczynska A, Bushell M. The complexity of miRNA-mediated repression. *Cell Death Differ*. 2015;22(1):22-33. doi:10.1038/cdd.2014.112.
259. Meckler JF, Bhakta MS, Kim MS, et al. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res*. 2013;41(7):4118-4128. doi:10.1093/nar/gkt085.
260. Delacôte F, Perez C, Guyot V, et al. High Frequency Targeted Mutagenesis Using Engineered Endonucleases and DNA-End Processing Enzymes. *PLoS One*. 2013;8(1):e53217. doi:10.1371/journal.pone.0053217.
261. Andreani M, Testi M, Battarra M, et al. Relationship between mixed chimerism and rejection after bone marrow transplantation in thalassaemia. *Blood Transfus*. 2008;6(3):143-149. doi:10.2450/2008.0051-07.
262. Mohrin M, Bourke E, Alexander D, et al. Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell*. 2010;7(2):174-185. doi:10.1016/j.stem.2010.06.014.
263. Manjunath N, Wu H, Subramanya S, Shankar P. Lentiviral delivery of short hairpin RNAs. *Adv Drug Deliv Rev*. 2009;61(9):732-745. doi:10.1016/j.addr.2009.03.004.
264. Lin SL, Miller JD, Ying SY. Intronic microRNA (miRNA). *J Biomed Biotechnol*. 2006;2006(4):26818.

doi:10.1155/JBB/2006/26818.
265. Yin H, Kanasty RL, Eltoukhy AA, Vegas AJ, Dorkin JR, Anderson DG. Non-viral vectors for gene-based therapy. *Nat Rev Genet*. 2014;15(8):541-555. doi:10.1038/nrg3763.
266. Bergmann T, Ehrhardt ES and A. Progress and Problems with Viral Vectors for Delivery of Talens. *J Mol Genet Med*. 2014;8(1). doi:10.4172/1747-0862.1000096.
267. Govindan G, Ramalingam S. Programmable Site-Specific Nucleases for Targeted Genome Engineering in Higher Eukaryotes. *J Cell Physiol*. 2016;231(11):2380-2392. doi:10.1002/jcp.25367.